GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Secular changes in sexual and natural selection against deleterious genetic mutations in humans

Dissertation

for the award of the degree

Doctor rerum naturalium (Dr. rer. nat.)

of the Georg-August-University School of Science (GAUSS)

submitted by

Ruben Cem Arslan,

born in Aachen, Germany

Göttingen, 2017

Thesis Committee

Lars Penke

Biological Personality Psychology, University of Goettingen

Markus Jokela

Institute of Behavioral Sciences, University of Helsinki

Members of the Examination Board

Reviewer: Lars Penke

Biological Personality Psychology, University of Goettingen

Second Reviewer: Markus Jokela

Institute of Behavioral Sciences, University of Helsinki

Further members of the Examination Board:

Annekathrin Schacht

Affective Neuroscience and Psychophysiology, University of Goettingen

Stefan Schulz-Hardt

Economic and Social Psychology, University of Goettingen

Julia Ostner

Behavioral Ecology, University of Goettingen

York Hagmayer

Cognition and Decision Making, University of Goettingen

Date of the oral examination: 25.10.2017

Danksagung

Dieses Dissertationsprojekt wäre nie zu einem Ende gekommen, ohne die vielen unterstützenden und willkommen ablenkenden Menschen in meinem Leben.
Mein Dank gebührt:

Meinem Doktorvater, Lars Penke. Sein Artikel über die evolutionäre Genetik der Persönlichkeit hat mein Interesse an der evolutionären Persönlichkeitsforschung geweckt, ein Praktikum bei ihm hat es vertieft und in nun vier Jahren Zusammenarbeit, vielen langen Gesprächen und einigen spätabendlichen Pizzen in Göttingen konnte ich hoffentlich ein paar weitere Antworten und Fragen beisteuern.

Markus Jokela, whose research I love to teach, because it shows how fun and clever personality research can be, and who agreed to thesis committee meetings in odd places under odd circumstances.

Roos Hutteman, die als meine entfernteste Cousine dafür gesorgt hat, dass ich mich in der Wissenschaft wohlfühle, der ich Fragen zu dieser fremden Welt stellen konnte, die maßgeblich mein Methodenwissen zu verschulden hat, mir regelmäßig geholfen hat und deren Kommentare auf meine Manuskripte unersetzlich waren.

Jaap Denissen, der mir als Hilfskraft in Berlin meine ersten Einblicke in die Wissenschaft ermöglicht hat, mich stets gefördert hat und durch den ich viele Menschen in dieser Danksagung erst kennengelernt haben.

Tanja Gerlach, meine Koautorin und Kollegin, die mir meine Inkompetenzen verzeiht, seit dem ersten Tag als Hilfskraft in Berlin bis zur Promotion und hoffentlich darüber hinaus.

Cyril Tata, who saved me from formr, so that working on it together is joyful again. Congratulations on your own replication project!

Anne Schacht, Stefan Schulz-Hardt, Julia Ostner und York Hagmayer, für ihre kurzfristige Bereitschaft Mitglieder meines Prüfungskomitees zu werden.

Paul-Christian Bürkner, der mir so viele Fragen zu Statistik beantwortet hat, dass ich in manchen Wochen mehr Worte mit ihm gewechselt habe als mit meiner Freundin.

Michael Dufner, dem einzigen Menschen, der mich im Büro anruft, mit dem ich Konferenznächte und Revisionen durchgebracht habe.

Meine Doktorandenkollegen, Tobias Kordsmeyer, Christoph von Borell, Julia Jünger, Tessa Cappelle, die das Joch mit mir getragen haben, sich mit mir über die Eigenarten dieses Berufs ausgelassen, notwendiges Eis gegessen, und Wunden geschrubbt haben.

Steve Gangestad, who, in many conversations, visits, and emails changed my mind and my drafts for the better.

Katharina Schilling, ohne die die Zyklusstudie nicht so erfolgreich gewesen wäre, und deren Abwesenheit aus Göttingen ich bedauere.

Kai Willführ, für rasante Diskussionen über Wissenschaft, Politik und Jagd, mit nicht immer klarer Thementrennung.

My many co-authors, who have supplied data, advice, their wits to make this work a hopefully good advertisement for the value of collaborative science.

Unseren studentischen Mitarbeitern, insbesondere Aileen Marske, Silvi Bradatsch, Selina Reinhard, Laura Botzet und Yasmin Thies, die mir viel abgenommen haben.

Meinen Studenten, an denen ich erklären üben konnte und durfte.

Meinen Göttinger Freunden, Philip, Shama, Filippo, Gitanjali und Audra, die mich ab und zu aus dem Büro und raus ins Kino, Bars, oder an den Esstisch gelockt haben.

Meinem Berliner Freundeskreis und teils baldigen Nachbarn, Stephan, Heidi, Manuel, David, Maike, Frauke, Nimrod, Katharina, Christian, Luca, Karl, Zooey, Christine, Isabelle und Maren. Ohne euch hätte mir diese Zeit nicht so viel Spaß gemacht. Mit euch möchte ich noch viele Flüsse flößen, Mahle genießen, Feste besuchen und organisieren und Abende verbringen.

Insbesondere Isabelle Habedank und Maren Fußwinkel, die mir in den letzten harten Wochen mit ihren Kommentaren und ihrer Freundschaft geholfen haben.

Meinen verstreuten Freunden Paul, Tobias, Niklas, Elena, Marin, Fredrik, Josh, Anne, die mich beherbergt haben, und sich meine Forschung erklären lassen haben.

Meinen Kobloggern, Anne Scheel, Julia Rohrer und Malte Elson, die mir während des letzten Jahren Freunde geworden sind, deren Gesellschaft und Humor mich einige Probleme in der Wissenschaft aushalten und andere verändern lassen.

Scott Alexander, whose writing inspires me to write, to think and to act better.

Andrew Gelman, whose blog taught me much of what I know about statistics and good scientific practice, and more than I care to know about basketball, whose voice has become a little person in my head.

Meiner (Wahl)familie, Grete, Nihan, Nedret, Andreas, Maggi, Norbert, Sebastian, Simon, Samuel, für warme Mahlzeiten, Unterstützung und unvorhersehbare Attacken auf mein Leibeswohl, die meine Reaktionszeiten schnell halten.

Meinem Vater, Necati Arslan, dessen Musik mich durch die Zeit, in der mir oft die Muße fehlte, begleitet hat und mir den Wert von Gelassenheit vermittelt hat.

Meiner Schwester, Rosa Arslan, die viele meiner Fehler dadurch verkörpert, dass sie sie nicht hat, von deren Warmherzigkeit ich zehre und lerne, die mich unterstützt und auf die Verlass ist.

Meiner Mutter, Ulla Arslan, die die Ergebnisse meiner zweiten Studie introspektiv schon vorher ermittelt hatte, vor der ich die Bedeutsamkeit meiner Forschung rechtfertigen muss, was meine Arbeit besser macht, aber vor allem als Vorbild für mich in allen Lebenslagen, durch Gerechtigkeit, Humor, Genuss und Bildung, was mich besser macht.

Meiner Verlobten, Sarah J. Lennartz. Deine Hilfe, Dein Humor, Deine Geduld, Dein kritischer Geist und Deine Liebe haben mich zu dem geformt, der ich heute bin. Ohne Dich hätte ich das nicht vermutlich geschafft und sicher nicht gewollt, denn mein Hauptantrieb ist Dich beeindrucken und Deiner Hochachtung gerecht werden zu wollen.

Table of Contents

1. Preface

This dissertation is not a cumulative, publication-based dissertation, but follows it in form. It includes three manuscripts, two of which have been accepted for publication, and one of which is under review.

Arslan, R. C., & Penke, L. (2015). Evolutionary Genetics. In *The Handbook of Evolutionary Psychology* (Vol. 2, pp. 1047–1066). New York: Wiley.

Arslan, R. C., Willführ, K. P., Frans, E., Verweij, K. J. H., Myrskylä, M., Voland, E., … Penke, L. (in press). Older fathers' children have lower evolutionary fitness across four centuries and in four populations. *Proceedings of the Royal Society B: Biological Sciences*.

Arslan, R. C., Schilling, K. M., Gerlach, T. M., & Penke, L. (in prep.). Using 26 thousand diary entries to show ovulatory changes in sexual desire and behaviour.

2. Introduction

All genetic variation once arose by mutation. Genetic variation is the substance on which natural selection can act, but mutations are random. Therefore, new variations usually have negative or no effects on an organism's evolutionary fitness (Keightley, 2012). The exceptions to this rule, for instance the heavily selected mutations causing lactase persistence, the ability to digest milk in adulthood (Tishkoff et al., 2007), are what drives the evolution of adaptations by natural selection.

This dissertation is *not* about these exciting exceptions, but about harmful mutations. We know that the average human child is born with 70-100 genetic variants that its parents did not carry (Kong et al., 2012; Rahbari et al., 2015; Ségurel, Wyman, & Przeworski, 2014). On average, around 2.2 of these are estimated to be harmful (Eyre-Walker & Keightley, 2007; Keightley, 2012). Why don't deleterious mutations build up irreversibly, each new generation carrying their new ones and the ones they inherited from their parents (Lesecque, Keightley, & Eyre-Walker, 2012)? One possible answer to this question seems to be sex, or more specifically recombination. Clonally reproducing organisms must suffer one genetic death (failure to reproduce) per mutation to remove it from the species' genetic pool, but with recombination some shuffled genomes carry many deleterious mutations and some few. This way, natural selection can purge deleterious mutations more efficiently. This benefit has long been thought to partially explain why organisms pay the twofold cost of sex, i.e. why they go through the trouble of having to find a mate and then only passing on half of their genes (Hartfield & Keightley, 2012; Kondrashov, 1988). But if recombination is sufficient, why shuffle your genome with someone else? One idea is that while natural selection is blind, favouring no particular direction, sexual mates can be very discerning. Through the preferences of

sexual mates, selection against deleterious mutations could accelerate (Andersson & Iwasa, 1996; Whitlock & Agrawal, 2009). Choosy mates might even be helped in their choice by handicaps, such as the peacock's tail, that only individuals in good condition can afford to impose on themselves (Zahavi, 1975).

Mutations might seem like a topic of little importance to psychology. They become interesting once we ask why studies of twins, families, and similarity at molecular genetic loci consistently show that individual differences in personality and intelligence are heritable (Lo et al., 2016; Penke, Denissen, & Miller, 2007; Plomin & Deary, 2014). Given that we also know that differences in personality and intelligence influence important life outcomes, including survival and reproduction (Jokela, 2012; Penke & Jokela, 2016), should not selection tend to remove these differences and fix them at whatever level is optimal for producing the most children? Again, mutations are one piece in this puzzle. Whereas selection moves a trait closer to the optimal level, most mutations will move the trait further away from the optimum. This continuous struggle is called mutation-selection balance and it is thought to play a major part in explaining why individual differences that are linked to evolutionary fitness, i.e. survival and reproduction, persist (Olson, 2012). Under this perspective, individuals with less adaptive trait values are expected to have higher *mutation load*, *i.e.* carry more harmful mutations (Penke et al., 2007).

In this work, I ask whether the balance between mutation and selection is fragile and easily upset. To me, the potential fragility of this balance appears to be a theme common to several worries that have been voiced in the literature (Alvergne & Lummaa, 2010; Crow, 1997; Lynch, 2016; Sartorius & Nieschlag, 2010).

Three worries

Delayed reproduction

One worry concerns the increasing age at which individuals tend to reproduce. The mother's, but even more so the father's age at birth are linked to the number of mutations that a child carries (Wong et al., 2016), leading to visible increases in genetic disease with advancing paternal age (D'Onofrio et al., 2014; Glaser & Jabs, 2004). The last forty years have seen an increase in the age at *first* birth for mothers, while data on ages at *all* births for both parents is hard to come by (Sartorius & Nieschlag, 2010). The worry that Crow (1997), and Sartorius and Nieschlag (2010) voiced is that unprecedented, increasingly delayed reproduction will lead to an onslaught of new mutations that deteriorates the average human condition. In the words of (Crow, 1997) "the greatest mutational health hazard in the human population at present is fertile old males". I will call this "the delay worry".

Relaxed selection

A second worry concerns the perception that natural selection is relaxed in modern times in large parts of the world (Crabtree, 2012; Crow, 1997, 2000; Keightley, 2012; Kondrashov, 1988; Lynch, 2016). Infant mortality is at an all-time low, most people have a similar number of children, and societal institutions like insurance, modern medicine and welfare mitigate problems that individuals and families would have previously faced on their own (Lynch, 2016). However, these factors do not necessarily imply that selection against deleterious mutations is relaxed. Some variation in survival and fitness may have been largely random so that decreases in mortality or less variation in reproductive success need not necessarily affect selection against genetic mutations. In other cases, selection may have relaxed at one stage of life, but tightened in another. For example, if children with a genetic disease survive thanks to modern medicine, they might still have trouble finding a

mate. In another example, although children will Down's syndrome can live long lives and rarely even reproduce, pregnancies are nowadays routinely screened for trisomy 21 and usually aborted in the case of detection (Mansfield, Hopfer, & Marteau, 1999). Last, relaxed selection is only a meaningful concept in reference to past times. For instance phenylketonuria (PKU) is a disease that causes brain damage, but can be prevented by removing phenylalanine from the diet. As long as the medical knowledge to diagnose and treat PKU remains, no large problems result from relaxed selection against it (National Institutes of Health Consensus Development Panel, 2001), as long as we are simply concerned with evolutionary fitness in the contemporary world. Still the worry could be justified from different perspectives, i.e. from a perspective of medical costs (Lynch, 2016), from a perspective on human qualities that values, for instance, intelligence for its societal benefits rather than its relationship to evolutionary fitness (Crabtree, 2012), or from a perspective concerned with the loss of civilizational knowledge and institutions through global catastrophes. In the words of (Kondrashov & Crow, 1993) "In human populations with a high living standard there is very little selection against minor deleterious mutations. However effective selection against them may have been in the past, it is not likely to operate efficiently now." I will call this the "the relaxation worry."

Altered mate choice

A third worry concerns not the strength, but the direction of selection. Some researchers are concerned that female mate preferences might have changed through the use of hormonal contraception (Alvergne & Lummaa, 2010). This worry arises from a broader literature in evolutionary psychology which posits that mate preferences may vary over the menstrual cycle in women (Gangestad & Thornhill,

2008). The idea is that natural selection would favour women who choose the best sexual partners and that this choice may differ adaptively depending on whether the sex is likely conceptive or non-conceptive. Most prominently, advocates of this perspective suggest women might have stronger preferences for males with *good genes* when conceptive. *Mutation load* is an important component of the concept of good genes, among other aspects such as interindividual genetic fit (e.g. compatibility on immune system genes), and adaptedness to the current environment (e.g. having high fat reserves in a society that frequently faces starvation). The latter two aspects are even more difficult to study than mutation load, because preferences for them might vary across women. Thus, many studies on *ovulatory shifts* have focused on male traits they believe to be indicative of low mutation load. These traits have included fluctuating asymmetry, masculinity, intelligence, and dominance (Gangestad, Thornhill, & Garver-Apgar, 2015; Gildersleeve, Haselton, & Fales, 2014), but perhaps most straightforwardly, research has also examined attractiveness. From an evolutionary point of view, the attractiveness of a male for a short-term sexual relationship (e.g. a one-night stand) should depend only on the genetic material potentially transferred during such an encounter. So, sexual selection against mutations (Whitlock & Agrawal, 2009) might work in part by shifting mate choice in the conceptive phase of the cycle to more strongly prefer sexually attractive men, especially for extra-pair copulations (Andersson & Iwasa, 1996). Because hormonal contraceptives inhibit the hormonal changes happening around ovulation midcycle, (Alvergne & Lummaa, 2010) have proposed that they might also affect mate preferences and choices to make them more similar to mate preferences in the luteal, non-conceptive part of the cycle, presumably decreasing the preference for sexually attractive men. Apart from the

pragmatic worry that women might end up unhappy with their partner if they go on or off hormonal contraception during a relationship, there is also an evolutionary worry. Ovulatory changes in mate choice might serve an important adaptive function, namely selection for *good genes*, that is lost. In the words of Alverge and Lummaa (2010) “pill might also have a non-negligible impact on mating decisions and subsequent reproduction. If this is the case, pill use will have implications for both current and future generations.” I will call this the “the mate choice worry”.

These three worries all relate in some way to upsets in the balance of mutation and selection. But are they justified? Are we hurtling towards a future, in which humans carry more deleterious mutations?

I cannot conclusively answer these questions, but I can draw on two different empirical studies that form the backbone of this dissertation (Arslan et al., in press; Arslan, Schilling, Gerlach, & Penke, in prep.), another study, to which I contributed but which is not part of this dissertation (Hill et al., 2017), and the existing scientific literature on the topic, some of which I reviewed in a book chapter on evolutionary genetics (Arslan & Penke, 2015).

Two empirical approaches

My first approach makes use of stochastically known genetic associations between relatives, similarly to twin and sibling studies. Namely, I made use of the strong relationship between the father’s age at offspring conception and offspring mutation load (Kong et al., 2012; Ségurel et al., 2014). Starting in puberty, spermatogonial stem cells keep dividing every 16 days. Whereas the entire pool of female germ cells, oocytes, has formed before the birth of a future mother, future fathers keep producing sperm. Continuous copying begets copying errors. Thus, fathers are

thought to be a main source of replication-driven mutations, although a less pronounced effect of the mother's age on mutations is also detectable (Wong et al., 2016).

Because maternal and paternal age are strongly correlated, paternal age on its own can explain a substantial proportion of the non-random variation in the number of *de novo* (as opposed to inherited) single nucleotide mutations – in one study almost all of it (Kong et al., 2012). This strong relationship makes it feasible to use paternal age as a *proxy* or placeholder variable for de novo mutations. Modern molecular genetics makes it possible to directly count the number of de novo mutations by sequencing the genomes of parents and their children and aligning them against each other (Deciphering Developmental Disorders Study, 2017). But sequencing is still expensive for larger samples and cannot easily be applied to historical datasets so that the indirect approach via paternal age is worthwhile.

My second approach stems from a different subdiscipline, evolutionary psychology. Here, I ask whether I can detect changes in preferences for mating partners in the middle of the menstrual cycle, when the probability of conception is highest. Theoretically my two approaches are connected through the concept of mutation load, but my second approach makes different assumptions and implements a very different methodology, complementing the first approach. Namely, the *good genes ovulatory shift hypothesis* predicts that women will change their mate choices around ovulation to obtain the best possible genes for their children, while they may have different preferences when non-conceptive, for instance obtaining a committed partner, who protects and provides resources (Gangestad & Thornhill, 2008; Gangestad et al., 2015).

Potentially affected psychological traits

In this work, I will discuss how traits of interest to psychologists might be affected by mutations, focusing on mental health, intelligence, and personality. Worries about secular changes in the balance between mutation and selection have been voiced (Alvergne & Lummaa, 2010; Crabtree, 2012; Lynch, 2016; Sartorius & Nieschlag, 2010) mainly because of the expected consequences for potentially affected traits such as intelligence, personality, and psychiatric disease. Hence, I will briefly review debates around secular trends in these traits. Ever since we have started keeping records, researchers have documented overall increases in intelligence (Flynn, 1987), economically valuable personality traits (Jokela, Pekkarinen, Sarvimäki, Terviö, & Uusitalo, 2017), but also for instance autism (Lundström, Reichenberg, Anckarsäter, Lichtenstein, & Gillberg, 2015; Wing & Potter, 2002). For some part, these increases are thought to merely reflect changes in how we measure and diagnose these traits and in how people respond to our tests (Lundström et al., 2015; Pietschnig & Voracek, 2015). Most of the remaining change is usually attributed to environmental causes, because the rate of change is too high for known evolutionary processes (Jokela et al., 2017; Pietschnig & Voracek, 2015). However, changes on the genetic level could still be taking place, albeit more slowly, and not necessarily in the same direction. In one example, an Icelandic study found that the average genetic propensity to complete higher education (predicted using results from genome-wide association studies) decreased by a small amount, probably through smaller families, at the same time as the average phenotypic level of education increased by much more, through political reforms (Kong et al., 2017).

My work's primary focus has been on psychological variation in the normal range, especially in intelligence and personality. Yet, answering my questions about the role

of mutations required that I broaden my perspective to individual differences outside the usual focus of psychologists. I examined data on survival, marriage, and reproductive success in my work on paternal age effects and data on attractiveness in my work on ovulatory changes. These traits have the advantage that evolutionary genetic understanding of their function is more advanced. They are all understood to directly contribute to evolutionary fitness or to constitute it. By contrast, my work on this dissertation was motivated in part by our failure to find paternal age effects on intelligence and personality (Arslan, Penke, Johnson, Iacono, & McGue, 2014). In this work, we used a sample of 1898 twin pairs to test whether we would observe negative associations between paternal age and offspring intelligence and personality after adjusting for the parents' intelligence and personality. We did not find any significant associations after adjustments, but we could not straightforwardly interpret this result. Did we fail because our sample size was too small (but other studies with more than 500 thousand siblings also found no effect; (Myrskylä, Silventoinen, Tynelius, & Rasmussen, 2013)? Or did we fail because the paternal age approach is too indirect? Or did we fail to find an association because there is none, because intelligence and personality are not sensitive to mutations?

Our original aim was to test the prediction made by Penke, Denissen, and Miller (2007) that intelligence, but not personality, would be found to be under mutation-selection-balance and hence be negatively affected by increased paternal age. But to do so, we first needed to test whether the paternal age effect approach is a viable way to assess the effects of mutations at all.

To this end, we needed outcomes that were clearly related to evolutionary fitness. And what could be closer to fitness than survival and reproductive success?

Because survival and pedigrees have been recorded for longer than intelligence and personality, these data were also suitable to examine the relaxation worry. If certain mutationally caused diseases and traits are less strongly selected against in modern times, the association between paternal age, an indicator of mutations, and fitness, a measure of selection, should relax in modern populations compared to older ones. At the same time, data on entire populations in different times and locations allowed me to assess the delay worry, namely whether reproduction is increasingly delayed, potentially leading to an unprecedented influx of new mutations.

3. Summary of Manuscript 1

In this book chapter, we wanted to popularise evolutionary genetics methods with evolutionary psychologists. Evolutionary psychology (EP) shares a meta-theory with behaviour genetics (BG), but for historical reasons many EP studies regard individual differences as little more than noise (Tooby & Cosmides, 1990). Evolutionary psychology has tended to focus on universal human monomorphic and sexually dimorphic adaptations. Meanwhile, BG has shown that individual differences are heritable using twin and family studies, but more recently also using molecular genetic work (Plomin & Deary, 2014; Turkheimer, 2000). However, BG has tended to be more data-driven and did not have a strong focus on how evolution maintains heritable variation. Consequently, although we have fairly precise numbers for the percentage of a trait that can be explained by genetic differences, we often have little idea why these genetic differences persist (Barton & Keightley, 2002).

In the chapter, we introduced the forces of mutation, selection, drift, and migration and how they can balance each other out. We addressed the common position that traits like intelligence and personality are selectively neutral, i.e. that the

balancing forces of mutation and drift maintain the genetic variation therein. Because intelligence and personality have been linked to evolutionarily important outcomes like mortality and fertility (Alvergne, Jokela, & Lummaa, 2010; Batty et al., 2009; Jokela, 2012; Kong et al., 2017; Penke & Jokela, 2016; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007), it seems unlikely that selection plays no role in their maintenance at all. According to (Penke et al., 2007) the most likely explanation for genetic variation in personality is some form of balancing selection, whereas mutation-selection-balance is more likely to explain intelligence differences. However, one central prediction from this theory, namely that genetic variants for personality would be easier to find, has since been falsified. Progress in genome-wide association studies for personality has been similarly slow as for intelligence (Davies et al., 2015; Lo et al., 2016). Hence, the role of mutation and selection is still to be determined for many psychological traits.

We then reviewed the evolutionary genetics toolkit. Often a classical behaviour genetic design, such as twin and family studies, is complemented by a molecular genetic design, such as genome-wide complex trait analysis, that serves a similar aim on the molecular level. Similarly, paternal age effect studies are complemented by sequencing parents and children and counting mutations and effects of inbreeding can be studied in the children of cousins or by measuring associations with runs of homozygosity.

Especially relevant for this dissertation is the question how selection can be studied directly. Although associations between mortality, fertility and psychological traits can be measured, there are shortcomings to this approach. First, measuring personality and intelligence prospectively, before mortality and lifetime reproductive

success are measured requires long follow-ups. Measuring personality and intelligence later in life always bears the risk of reverse causation, for instance maybe having children makes one more emotional (Jokela, Kivimäki, Elovainio, & Keltikangas-Järvinen, 2009). Second, associations observed in a contemporary population may not be invariant over time and place. Potentially, a trait that has been positively selected throughout most of human prehistory and history is negatively selected in the modern world, or vice versa.

One way around this might be to instead examine mate preferences. Although sexual and natural selection do not necessarily act in the same direction, sexual selection is important in its own right (Long, Agrawal, & Rowe, 2012). (Buss, 1989) has shown that mate preferences are relatively invariant across cultures, certainly more so than the average number of children. Perhaps mate preferences preserve ancient selection pressures even in the modern world.

4. Summary of Manuscript 2

In this manuscript, we examined paternal age effects on offspring fitness. As an index of fitness, we mainly focused on the offspring's number of children. Our goal was to isolate the mutational aspect of paternal age effects. To do so, we needed to rule out many alternative pathways in which paternal age might be associated with offspring fitness. Most importantly, we compared full siblings in a multilevel regression model and adjusted for the average paternal age within the family. Because all children of a couple have the same random chance to inherit some of their genes, this approach allowed us to rule out that less fit fathers simply found partners later in life and their children inherited low fitness. However, we also had to

adjust for confounders that still differed between siblings and correlated with paternal age, such as birth order and parental loss.

We examined four different populations. One was 20th-century Sweden, our data were based on governmental records. The other three populations were from pre-industrial times (1720-1850). Church records were digitalised and used to reconstitute genealogies for the Saint-Lawrence valley, Québec (Canada), the Krummhörn (Germany) and four historical Swedish regions.

From our analyses of these genealogies, three main conclusions are relevant to this dissertation.

First, we found negative paternal age effects in all four populations that we examined. They were small, as predicted, but remained after adjusting for a lot of potential confounds in our robustness checks.

Second, average paternal and maternal ages at birth rose in 20th-century Sweden from 1970 onwards. However, from 1930 until 1970 they dropped. In 2010, they were at similar levels as in 1930 (around 33). More interestingly though, average parental ages were still below historical averages of the three pre-industrial populations that we also examined. This seems counter-intuitive only because most previous studies focused on maternal age at first birth. Compared to a historical baseline, the age at first birth is indeed delayed, but because people are also having far fewer children on average, the age at last birth and the average age at birth are earlier.

Third, differences between paternal age effects in the pre-industrial populations and 20th-century Sweden were substantial for infant mortality but not for the aggregate

effect on reproductive success. Across populations, we found no replicable and robust effects on survival of the first 15 years or the odds of getting married, nor clear differences between populations. Although paternal age predicted increased infant mortality in all four populations, infant mortality on average is so much lower in 20th-century Sweden that the effect was insubstantial in comparison to the other populations. However, when examining the paternal age effect on aggregate offspring reproductive success (including low reproductive success caused by early mortality), two things became clear: 20th-century Sweden did not stand out as exhibiting the smallest effect size, the effect size in Québec was smaller. Moreover, across 26 different model specifications all of which had some degree of plausibility, the effect sizes varied more than across populations. Because we probably cannot identify one true, best model, we cannot clearly conclude that selection against mutations is relaxed.

We were interested in the question whether relaxed postnatal survival selection is compensated by sexual selection later in life. However, we only had data on marriage and divorces, which are poor indices of mating success in 20th-century Sweden, because marriage is no longer a social or legal prerequisite for being in a relationship, cohabiting, or having children. Further, because we had no data on abortions, we could not clearly conclude whether infant survival selection was truly relaxed or displaced to before birth. Approximately 20% of all pregnancies are aborted in the modern Western world, but only few of these abortions are "therapeutic", i.e. aim to end a pregnancy where a potential birth defect was detected. Still, because the paternal age effects we found were quite small, these abortions might explain (part of) the difference.

Summary of Manuscript 3

In this study, we collected daily online self-reports from a large sample of women. A final sample of 1043 women filled out a short survey every day until they had contributed up to 40 days.

Our goal was to replicate and extend previous studies' reports that women's sexual interests change around ovulation. To this end, we asked our participants about their menstruation dates and contraceptive methods in the study. From the menstruation dates, we could then estimate the probability of being in the fertile window for each diary day. In multilevel models, we then tested whether being in the fertile window was associated with psychological changes. We also tested whether that such changes were absent among hormonal contraception users, who do not experience ovulation and the concurrent hormonal changes.

A key theoretical prediction was that women's assessment of their partners' attractiveness for a short-term sexual relationship should moderate the shifts in sexual desire in such a fashion, that women with more attractive partners experience increases in in-pair desire, but not extra-pair desire and vice versa for women with less attractive partners. The purported evolutionary function of this moderation pattern is to obtain *good genes* for the offspring.

Previous studies had supported this prediction in small studies, but many methodological criticisms of the previous literature were raised. Namely, previous studies had often used small sample sizes, often gathered data from women only on one or two days and did not preregister their methodology. This combination of problems is now thought to lead to overestimation of effect sizes and false positives.

In our study, we wanted to prevent these problems by preregistering our approach and recording responses from a large sample of women over 40 days.

We found ovulatory changes, that is peaks in the fertile window restricted to non-hormonal contraceptive users, for several outcomes. Namely, we replicated changes in extra- and in-pair sexual desire and behaviour, and in self-perceived sexual desirability.

However, we did not confirm the predicted moderation patterns. Even though our sample size of naturally cycling women was larger than the combined sample sizes of previous studies and about ten times larger than the average previous study on the subject, we found no significant moderation patterns.

Previous studies had mostly excluded women using hormonal contraception from participating to save costs. Our online approach allowed us to include them in our study and directly test whether the ovulatory changes observed among naturally cycling women were absent. They were. Hence, it seems possible that hormonal contraception would flatten cyclical changes in mate preferences. However, we found no evidence for such changes when we examined the moderation of sexual desire changes by the partner's short-term attractiveness. We would have predicted hormonal contraception users to permanently have the sexual desire of naturally cycling women in the luteal phase.

Future studies should examine whether mate preferences change across the cycle at all, in studies that also include single women. If they do not, then hormonal contraception is unlikely to have any effect on mate preferences either, although randomised controlled trials are necessary to rule this out with finality.

On the basis of our results, it seems possible that hormonal contraception does affect female sexuality in a measurable way, by flattening variability across the menstrual cycle. Mean levels may also be affected, but experimental studies are necessary to test this because of confounding variables. In our study, women using hormonal contraception differed from non-users not only in contraceptive method but also in other ways, mainly in age. Unfortunately, existing randomised placebo-controlled trials of hormonal contraceptives usually ignore the menstrual cycle. Package leaflets for hormonal contraception currently point out changes in libido as potential side effects, but are very unspecific. The reason for this might be that there is large heterogeneity in how sexuality varies across the cycle when not taking hormonal contraception, and thus, response to it may also vary across individuals.

5. General Discussion

The balance between mutations and selection is probably one of the main reasons why genetic differences persist (Olson, 2012). However, linking this concept to psychological research is difficult, because many fundamental aspects are still debated. In this dissertation, I tested one rather straightforward prediction, namely that mutational paternal age effects on fitness would be robust and replicable across populations. I also tested a less straightforward prediction, which nevertheless played a large role in the evolutionary psychological literature, namely that women's in-pair and extra-pair sexual desire would change across the menstrual cycle and that their perceptions of their partners' sexual attractiveness would moderate these shifts.

Both of my approaches' results were consistent with sexual selection not playing a major role in the selection against deleterious mutations. In the paternal age effect

studies, I would have expected to find a decrease in the odds of marriage with advancing paternal age. In the cycle studies, I would have expected to see the predicted ovulatory shift moderation pattern. Does this rule out that sexual selection plays any role in the selection against mutations in humans? No. In the paternal age studies, marriage is a coarse measure of mating success, but the only one that I could obtain from genealogies. I did observe an effect on number of children, after adjusting for differences in survival, but cannot tell from the data whether this was driven by differences in fertility, prenatal mortality of offspring, or the quality of the obtained partners. Quantitatively, it seems likely that survival selection plays a bigger role in selection against mutations than sexual selection. In the cycle studies, the partner's sexual attractiveness may not be as good an indicator of mutation load as previous studies had assumed, although it certainly was a strong candidate. Perhaps more importantly, variation of mate preferences over the menstrual cycle to engender extra-pair copulations, is not a necessary feature of sexual selection. Although extra-pair offspring would increase the strength of sexual selection compared to a monogamous baseline (Andersson & Iwasa, 1996), many Western human populations are serially monogamous, allowing the continued operation of sexual selection (Courtiol, Pettay, Jokela, Rotkirch, & Lummaa, 2012).

Identifying a causal, mutational effect of paternal age on fitness in humans is a difficult task. Through robustness analyses and replication across populations, I tried to test how robustly the effect could be shown. Many published studies did not adjust for confounds to the same extent that we did. Still, sibling comparison designs seem like a worthwhile approach to the problem that avoids many important confounds. A newer paternal age effect study by (Carslake, Tynelius, van den Berg, Davey Smith, & Rasmussen, 2017), using very large (>1.6m) sibling comparison samples from

Sweden, show tiny negative paternal age effects on offspring intelligence of approximately 0.07 standard deviations per decade of paternal age. Their estimates are consistent with our estimates (Arslan et al., 2014), although we lacked the statistical power to rule out sampling error. Other studies (D'Onofrio et al., 2014; Frans, MacCabe, & Reichenberg, 2015) report strong associations with various psychiatric diseases, but (Carslake et al., 2017) caution that there are further methodological pitfalls that such research often overlooks and (Gratten et al., 2016) caution that many estimated effect sizes are too large to be plausibly explained by mutations. We tried to make sure to avoid these pitfalls by replicating across populations, comparing our estimates to population genetic parameters and using many robustness checks. Still, in conclusion, it seems as if molecular genetic techniques will ultimately prove to be the superior way to test whether mutation-selection balance maintains variation in psychological traits. In one such study, which is not part of this dissertation, we used genomic relatedness estimates based on genetic variants imputed to the Haplotype Reference Consortium (Haplotype Reference Consortium, 2016). By stratifying low-level genetic relatedness by the frequency of the minor allele, we could show that rare variants were disproportionately involved in intelligence genetics, hinting at mutation-selection balance. Analyses based on higher levels of relatedness, for instance between cousins and siblings, supported the same conclusion. For neuroticism and extraversion, the two methods were less consistent (Hill et al., 2017).

Given that some new research shows a relationship between traits like intelligence and mutation load, I have to ask if the worries about increases in mutation load are justified.

Three worries

A main conclusion of my book chapter on evolutionary genetics (manuscript 1) was that understanding the balancing mechanisms that maintain genetic variation in psychological traits is not only interesting as a basic research question, but also directly relates to societal trends. Changes in policy and mores influence demography, reproductive timing, and the direction and strength of selection. Are some of these changes worrisome from a perspective of mutation-selection balance?

Delayed reproduction

Parental ages have been increasing since 1970 in 20th-century Sweden. There is no debating this, but a) increases in maternal ages at first birth were not a good guide to the smaller increases in average parental age at birth b) with context from three pre-industrial populations and from Sweden in 1930, the average parental age in Sweden in 2009 does not appear unprecedentedly high, but well within the bounds of previously observed variation. Hence, the average replication-driven mutation load of our population is probably also not unprecedentedly high. Excepting some births to older mothers - impossible before in-vitro-fertilisation - children do not have much older parents than observed in some of our pre-industrial populations.

Relaxed selection

The second worry I introduced is that the strength of selection against mutations has relaxed in the modern world. (Lynch, 2016), who most fully formulated the argument, was especially concerned with medical advances, which reduce the harm caused by deleterious germ line and somatic mutations, but was also worried about the low variability in number of offspring (family size). Restriction of the variance in number of offspring is a common strategy to reduce the efficiency of selection obtain

mutation accumulation lines in *Drosophila*. Lynch (2016) mentions some but neglects other factors that can work in the opposite direction, for example the increased population size of humans, decreased inbreeding, increased mobility, and increased ability to exercise mate choice. All may boost the efficacy of selection (Gazave, Chang, Clark, & Keinan, 2013; Keightley, 2012; Reed & Aquadro, 2006). Most importantly, his arguments and my counter-arguments must be tested empirically, because we cannot be sure that no other factors were neglected.

Comparing paternal age effect sizes across populations is one such empirical test, but it cannot resolve this worry with any finality. Still, even though the effects on infant survival were diminished in 20th-century Sweden, effects on the number of children persisted. In addition, the effect on overall offspring fitness was descriptively smaller in Québec than in 20th-century Sweden. This does not mean selection is not relaxed at all. After all, the population of the Saint Lawrence valley in Québec, as a small founding population, may also have experienced diminished selection against mutations (Casals et al., 2013). Yet, the effect sizes in the 20th-century are not significantly different from those in the pre-industrial populations. This makes it less likely that an unprecedented mutation load is currently accumulating (Keightley, 2012).

There are also molecular genetic ways to study relaxed selection and average mutation load. Such studies (reviewed in Simons & Sella, 2016) have focused on differences in mutation load between populations with different demographic histories. Severe population bottlenecks can lead to an increase in the deleterious mutation load. Evidence from these studies converged with our results, because summary indices of deleterious load (e.g. number of nonsynonymous derived alleles)

were not significantly different between ten contemporary populations (from West African Yoruba to the French and Chinese Han; (Simons & Sella, 2016). This was the case, even though they differed in recent population history and presumably also the spread of modern medicine and social transfers. As far as I am aware, similar molecular genetic indices have not yet been used to test for changes in mutation load over recent periods in the same populations, but molecular genetic methods are probably not sufficiently powerful at present genome sequence sample sizes to detect the small expected changes over short periods.

Altered mate choice

We interpret our findings as showing that the psychological changes around ovulation that occur for naturally cycling women are suppressed completely by hormonal contraceptives. To a lesser extent the same holds true for the psychological changes around menstruation. If mate preferences and choices varied because of the hormonal changes surrounding ovulation, they would probably also be suppressed. However, our findings shed doubt on claimed mate preference variation around ovulation. In our study, several measures of the partner's attractiveness did not moderate the changes in sexual desire. Another recently published large study also challenges previous reports of ovulatory changes in preferences for masculinity (Jones et al., 2017). We think this is an important area for future research. Although preferences for masculinity and short-term attractiveness may not change across the cycle, other mate preferences might. Furthermore, if certain theoretical predictions in the literature are correct, ovulatory changes might be strongest in extra-pair desire (Gangestad et al., 2015). If the pill made women more monogamous by inhibiting an ovulatory increase in extra-pair desire, this would be an important side effect to know about, both for the user and for

understanding secular change in sexual selection. Most likely, ovulatory changes are not the same for every woman and neither are the effects of hormonal contraceptives. Our data show some initial evidence of these interindividual differences, but more research is needed to show that these differences are stable and can be measured reliably.

6. Conclusion

In this dissertation, I showed that research on mutation-selection-balance can answer exciting basic research questions while at the same time speaking to worries about societal and demographic trends. We found evidence that selection prevents a build-up of mutations, but sexual selection did not seem to play an especially important role in this. Moreover, we found that selection continues to act against mutations in 20th-century Sweden, but we could not rule out with certainty that it has relaxed slightly. We also found evidence running counter to the notion that hormonal contraception alters mate choice, and thus sways sexual selection, but we only examined one aspect of mate preferences.

Our research cannot fully allay worries about relaxed selection, delayed reproduction, and altered mate choices, but we reported evidence that, given proper context and comparisons, these changes do not seem drastic. The balance between mutation and selection may not be as fragile as some have predicted, but it is clearly a topic worth examining. We call for further careful and empirical examination of this topic and the many other factors that may affect strength and efficacy of selection.

7. References

Alvergne, A., Jokela, M., & Lummaa, V. (2010). Personality and reproductive success in a high-fertility human population. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 11745–11750. doi:10.1073/pnas.1001752107

Alvergne, A., & Lummaa, V. (2010). Does the contraceptive pill alter mate choice in humans? *Trends in Ecology & Evolution*, *25*, 171–179. doi:10.1016/j.tree.2009.08.003

Andersson, M., & Iwasa, Y. (1996). Sexual selection. *Trends in Ecology & Evolution*, *11*, 53–58. doi:10.1016/0169-5347(96)81042-1

Arslan, R. C., & Penke, L. (2015). Evolutionary Genetics. In *The Handbook of Evolutionary Psychology* (Vol. 2, pp. 1047–1066). New York: Wiley. doi:10.1002/9781119125563.evpsych245

Arslan, R. C., Penke, L., Johnson, W., Iacono, W. G., & McGue, M. (2014). The effect of paternal age on offspring intelligence and personality when controlling for parental trait levels. *PLoS ONE*, *9*, e90097. doi:10.1371/journal.pone.0090097

Arslan, R. C., Schilling, K. M., Gerlach, T. M., & Penke, L. (in prep.). Using 26 thousand diary entries to show ovulatory changes in sexual desire and behaviour.

Arslan, R. C., Willführ, K. P., Frans, E., Verweij, K. J. H., Myrskylä, M., Voland, E., … Penke, L. (in press). Older fathers' children have lower evolutionary fitness across four centuries and in four populations. *Proceedings of the Royal Society B: Biological Sciences*.

Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*, *3*, 11–21. doi:10.1038/nrg700

Batty, G. D., Wennerstad, K. M., Smith, G. D., Gunnell, D., Deary, I. J., Tynelius, P., & Rasmussen, F. (2009). IQ in early adulthood and mortality by middle age: Cohort study of 1 million Swedish men. *Epidemiology*, *20*, 100–109. doi:10.1097/EDE.0b013e31818ba076

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49. doi:10.1017/S0140525X00023992

Carslake, D., Tynelius, P., van den Berg, G., Davey Smith, G., & Rasmussen, F. (2017). Associations of parental age with health and social factors in adult offspring. Methodological pitfalls and possibilities. *Scientific Reports*, *7*, 45278. doi:10.1038/srep45278

Casals, F., Hodgkinson, A., Hussin, J., Idaghdour, Y., Bruat, V., de Maillard, T., … Awadalla, P. (2013). Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. *PLoS genetics*, *9*, e1003815. doi:10.1371/journal.pgen.1003815

Courtiol, A., Pettay, J. E., Jokela, M., Rotkirch, A., & Lummaa, V. (2012). Natural and sexual selection in a monogamous historical human population. *Proceedings of the National Academy of Sciences*, *109*, 8044–8049. doi:10.1073/pnas.1118174109

Crabtree, G. R. (2012). Our fragile intellect. Part I. *Trends in Genetics*, 1–3. doi:10.1016/j.tig.2012.10.002

Crow, J. F. (1997). The high spontaneous mutation rate: Is it a health risk? *Proceedings of the National Academy of Sciences*, *94*, 8380–8386. doi:10.1073/pnas.94.16.8380

Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics*, *1*, 40–47. doi:10.1038/35049558

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., … Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: A meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Molecular Psychiatry*, *20*, 183–192. doi:10.1038/mp.2014.188

Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, *542*, 433–438. doi:10.1038/nature21062

D'Onofrio, B. M., Rickert, M. E., Frans, E., Kuja-Halkola, R., Almqvist, C., Sjölander, A., … Lichtenstein, P. (2014). Paternal age at childbearing and offspring psychiatric and academic morbidity. *JAMA Psychiatry*, *71*, 432–438. doi:10.1001/jamapsychiatry.2013.4525

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, *8*, 610–618. doi:10.1038/nrg2146

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171–191. doi:10.1037/0033-2909.101.2.171

Frans, E., MacCabe, J. H., & Reichenberg, A. (2015). Advancing paternal age and psychiatric disorders. *World Psychiatry*, *14*, 91–93. doi:10.1002/wps.20190

Gangestad, S. W., & Thornhill, R. (2008). Human oestrus. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 991–1000. doi:10.1098/rspb.2007.1425

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2015). Women's sexual interests across the ovulatory cycle. *The Handbook of Evolutionary Psychology*. doi:10.1002/9781119125563.evpsych114

Gazave, E., Chang, D., Clark, A. G., & Keinan, A. (2013). Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics*, *195*, 969–978. doi:10.1534/genetics.113.153973

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. doi:10.1037/a0035438

Glaser, R. L., & Jabs, E. W. (2004). Dear old dad. *Science of Aging Knowledge Environment*, *2004*, 1–11. doi:10.1126/sageke.2004.3.re1

Gratten, J., Wray, N. R., Peyrot, W. J., McGrath, J. J., Visscher, P. M., & Goddard, M. E. (2016). Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. *Nature Genetics*, *48*, 718–724. doi:10.1038/ng.3577

Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*, *48*, 1279–1283. doi:10.1038/ng.3643

Hartfield, M., & Keightley, P. D. (2012). Current hypotheses for the evolution of sex and recombination. *Integrative zoology*, *7*, 192–209. doi:10.1111/j.1749-4877.2012.00284.x

Hill, W. D., Arslan, R. C., Xia, C., Luciano, M., Amador, C., Navarro, P., … Penke, L. (2017). Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *bioRxiv*. doi:10.1101/106203

Jokela, M. (2012). Birth-cohort effects in the association between personality and fertility. *Psychological Science*, *23*, 835–841. doi:10.1177/0956797612439067

Jokela, M., Kivimäki, M., Elovainio, M., & Keltikangas-Järvinen, L. (2009). Personality and having children: A two-way relationship. *Journal of Personality and Social Psychology*, *96*, 218–230. doi:10.1037/a0014058

Jokela, M., Pekkarinen, T., Sarvimäki, M., Terviö, M., & Uusitalo, R. (2017). Secular rise in economically valuable personality traits. *Proceedings of the National Academy of Sciences*, *114*, 201609994. doi:10.1073/pnas.1609994114

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., … DeBruine, L. M. (2017). Women's preferences for facial masculinity are not related to their hormonal status. *bioRxiv*, 136549. doi:10.1101/136549

Keightley, P. D. (2012). Rates and fitness consequences of new mutations in humans. *Genetics*, *190*, 295–304. doi:10.1534/genetics.111.134668

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, *336*, 435–440. doi:10.1038/336435a0

Kondrashov, A. S., & Crow, J. F. (1993). A molecular approach to estimating the human deleterious mutation rate. *Human Mutation*, *2*, 229–234. doi:10.1002/humu.1380020312

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, *488*, 471–475. doi:10.1038/nature11396

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., … Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, e727–e732. doi:10.1073/pnas.1612113114

Lesecque, Y., Keightley, P. D., & Eyre-Walker, A. (2012). A Resolution of the Mutation Load Paradox in Humans. *Genetics*, *191*, 1321–1330. doi:10.1534/genetics.112.140343

Lo, M.-T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C.-C., Wang, Y., … Chen, C.-H. (2016). Genome-wide analyses for personality traits identify six genomic loci

and show correlations with psychiatric disorders. *Nature Genetics*, *49*, 152–156. doi:10.1038/ng.3736

Long, T. A. F., Agrawal, A. F., & Rowe, L. (2012). The effect of sexual selection on offspring fitness depends on the nature of genetic variation. *Current Biology*, *22*, 204–208. doi:10.1016/j.cub.2011.12.020

Lundström, S., Reichenberg, A., Anckarsäter, H., Lichtenstein, P., & Gillberg, C. (2015). Autism phenotype versus registered diagnosis in Swedish children: Prevalence trends over 10 years in general population samples. *British Medical Journal*, *350*, h1961. doi:10.1136/bmj.h1961

Lynch, M. (2016). Mutation and human exceptionalism: Our future genetic load. *Genetics*, *202*, 869–875. doi:10.1534/genetics.115.180471

Mansfield, C., Hopfer, S., & Marteau, T. M. (1999). Termination rates after prenatal diagnosis of Down syndrome, spina bifida, anencephaly, and Turner and Klinefelter syndromes: A systematic literature review. *Prenatal Diagnosis*, *19*, 808–812. doi:10.1002/(SICI)1097-0223(199909)19:9<808::AID-PD637>3.0.CO;2-B

Myrskylä, M., Silventoinen, K., Tynelius, P., & Rasmussen, F. (2013). Is later better or worse? Association of advanced parental age with offspring cognitive ability among half a million young Swedish men. *American Journal of Epidemiology*, *177*, 649–655. doi:10.1093/aje/kws237

National Institutes of Health Consensus Development Panel (2001). National institutes of health consensus development conference statement: phenylketonuria: Screening and management, October 16–18, 2000. *Pediatrics*, *108*, 972–982. doi:10.1542/peds.108.4.972

Olson, M. V. (2012). Human genetic individuality. *Annual Review of Genomics and Human Genetics*, *13*, 1–27. doi:10.1146/annurev-genom-090711-163825

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality*, *21*, 549–587. doi:10.1002/per.629

Penke, L., & Jokela, M. (2016). The evolutionary genetics of personality revisited. *Current Opinion in Psychology*, *7*, 104–109. doi:10.1016/j.copsyc.2015.08.021

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science*, *10*, 282–306. doi:10.1177/1745691615577701

Plomin, R., & Deary, I. J. (2014). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry*, *20*, 98–108. doi:10.1038/mp.2014.105

Rahbari, R., Wuster, A., Lindsay, S. J., Hardwick, R. J., Alexandrov, L. B., Al Turki, S., … Hurles, M. E. (2015). Timing, rates and spectra of human germline mutation. *Nature Genetics*, *48*, 126–133. doi:10.1038/ng.3469

Reed, F. A., & Aquadro, C. F. (2006). Mutation, selection and the future of human evolution. *Trends in Genetics*, *22*, 479–484. doi:10.1016/j.tig.2006.07.005

Roberts, B. W. B., Kuncel, N., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313–345. doi:10.1111/j.1745-6916.2007.00047.x

Sartorius, G. A., & Nieschlag, E. (2010). Paternal age and reproduction. *Human Reproduction Update*, *16*, 65–79. doi:10.1093/humupd/dmp027

Ségurel, L., Wyman, M. J., & Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. *Annual Review of Genomics and Human Genetics*, *15*, 47–70. doi:10.1146/annurev-genom-031714-125740

Simons, Y. B., & Sella, G. (2016). The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Current Opinion in Genetics & Development*, *41*, 150–158. doi:10.1016/j.gde.2016.09.006

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., … Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, *39*, 31–40. doi:10.1038/ng1946

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, *58*, 17–67. doi:10.1111/j.1467-6494.1990.tb00907.x

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*, 160–164. doi:10.1111/1467-8721.00084

Whitlock, M. C., & Agrawal, A. F. (2009). Purging the genome with sexual selection: Reducing mutation load through selection on males. *Evolution*, *63*, 569–582. doi:10.1111/j.1558-5646.2008.00558.x

Wing, L., & Potter, D. (2002). The epidemiology of autistic spectrum disorders: Is the prevalence rising? *Mental Retardation and Developmental Disabilities Research Reviews*, *8*, 151–161. doi:10.1002/mrdd.10029

Wong, W. S. W., Solomon, B. D., Bodian, D. L., Kothiyal, P., Eley, G., Huddleston, K. C., … Niederhuber, J. E. (2016). New observations on maternal age effect on

germline de novo mutations. *Nature Communications*, *7*, 10486. doi:10.1038/ncomms10486

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214. doi:10.1016/0022-5193(75)90111-3

Appendix A.

Manuscript 1 (Evolutionary Genetics)

Chapter 45 Evolutionary Genetics

Ruben C. Arslan and Lars Penke

Introduction

When Charles Darwin developed the theory of evolution, he knew nothing about genetics. Hence, one of its biggest weaknesses was that Darwin had to base it on crude ideas of inheritance. Around the same time, Gregor Mendel discovered the laws of inheritance, but the scientific community initially failed to appreciate his work's importance. It was only in the 1930s that Dobzhansky, Fisher, Haldane, Wright, Mayr and others unified genetics and the theory of evolution in the "modern synthesis." Still, the modern synthesis was built on a basic understanding of genetics, with genes merely being particulate inherited information. The basics of molecular genetics, like the structure of DNA, were not discovered until the 1950s. When modern evolutionary psychology emerged from ethology and sociobiology in the late 1980s, it had a strong emphasis on human universals, borne from both the assumption that complex adaptations are monomorphic (or sexually dimorphic) and have to go back to at least the last common ancestor of all humans, and the methodological proximity to experimental cognitive psychology, which tends to treat individual differences as statistical noise. As a consequence, genetic differences between people were marginalized in evolutionary psychology (Tooby & Cosmides, 1990). Evolutionary psychology and behavior genetics developed nearly orthogonally for over a decade. Behavior geneticists discovered that virtually every psychological or behavioral difference shows genetic variation (Turkheimer, 2000) and that the molecular genetic underpinnings of most heritable traits are far more complex than assumed in the modern synthesis. Meanwhile, evolutionary psychologists increasingly realized the importance of genetic variation, for example, in models of sexual selection for attractiveness, intelligence, and other assumed honest signals of genetic quality (Gangestad & Simpson, 2000) or heritable variation in life history traits (see Miller & Penke, 2007). During the past decade, evolutionary genetics gradually gained acceptance among evolutionary psychologists (Buss & Hawley, 2011; Gangestad & Yeo, 1997; Buss & Penke, 2014; Penke, Denissen, & Miller, 2007), though most still defer fully incorporating the genetic perspective (Miller, 2011).

Evolutionary genetics is concerned with the *mechanisms* that explain the existence and maintenance of genetic variation in traits. All else equal, one would expect selection to deplete genetic variation in heritable traits related to

fitness eventually (Penke et al., 2007). However, such genetic variation is ubiquitous and underlies stable individual differences that play prominent roles in psychological theories, be it as traits under intersexual (e.g., attractiveness, agreeableness, intelligence; Buss, 1989) and intrasexual selection (masculinity, aggressiveness; Puts, this volume), life history traits, formidability in recalibration theory (Sell, Tooby, & Cosmides, 2009), sociometer sensitivity (Denissen & Penke, 2008), perceived vulnerability to infection in the behavioral immune system (Schaller & Park, 2011), attachment security (Rholes & Simpson, 2006), or the tendency to show strong reciprocity in cooperation (Fehr, Fischbacher, & Gächter, 2002). Though these theories ascribe adaptive roles to individual differences, more or less explicitly linking them to fitness, their genetic variation is often taken for granted.

Evolutionary genetics can help evolutionary psychologists unearth clues to the ultimate reasons behind, for example, humans' cognitive faculties that go beyond what can gleaned through paleontology and archaeology (Enard, Messer, & Petrov, 2014). This information can have very practical implications, such as helping to understand how natural and sexual selection, when altered through changing mores or policy, will affect certain traits.

One aim of this chapter is thus to introduce some of the tools available to researchers in evolutionary genetics. Prior to that, we provide an overview of the forces of evolution and how their interactions can maintain genetic variation. To illustrate the various ways in which evolution can maintain individual differences, we will often invoke specific traits that seem to serve as good, didactically useful examples. The general approach, however, would be applicable to all sorts of traits, including those with relevance to evolutionary psychological theories. Rarely have all possible explanations been weighed explicitly in the literature; we thus tried to refrain from definite statements. With this caveat in mind, we believe that our examples will help evolutionary psychologists make use of the rich theoretical framework that evolutionary genetics provides.

Genetic Architecture

Some research in molecular genetics has been carried out with the aim of characterizing the *genetic architecture* of traits, sometimes also called the genotype-phenotype map (Mackay, 2001). The genetic architecture of a trait can provide important clues to the evolutionary history and the mechanisms that govern the maintenance of genetic variation in the trait (Penke et al., 2007). Characterizing the genetic architecture of a quantitative trait would ideally involve its robustness to mutations (*canalization*) as well as its *evolvability*. It would also imply gauging its degree of *pleiotropy* (whether the genes involved also have simultaneous other effects) and the importance of *nonadditive genetic variation* (i.e., epistasis and dominance, variation that does not breed true to the next generation). Unfortunately, many examinations of the genetic architecture are limited to estimates of the number and effect size of

involved genetic variants. Often the goal in such examinations is predicting which molecular genetic studies will succeed in the gene hunt and lead to biological pathways and drug targets, not to discover the ultimate, evolutionary explanations for heritable variation in a trait. In this chapter, we hope to suggest conceptual approaches to the latter goal.

It may feel like a step back from identifying causative genetic variants, but we feel it is prudent to set aside the exciting prospects of what a successful gene hunt might entail (Chabris et al., 2013) and the different ideas about how we might succeed at that (Graur et al., 2013; Mitchell, 2012), focusing instead on finding common theoretical ground.

Researchers disagree how, if ever, we might explain a substantial portion of the "missing heritability" (Mitchell, 2012), the observable genetic variation left unexplained by molecularly identified genetic variants. The limits of currently available tools can sometimes act as blinders, so that some theoretically plausible genetic architectures are hidden in our blind spots. Fortunately, as rapid technological and statistical development in molecular genetics adds to our tool kit, fewer blind spots should impede us. Humility is still very appropriate, though, considering fairly principal problems such as the sheer parameter explosion that is encountered when relating genomic sequences to traits (but see Ma, Clark, & Keinan, 2013).

Neither should we be too eager to jump to the conclusion that our purported core traits will be reflected at the genetic level. For example, Mitchell (2012) argued against the continuous liability-threshold model of psychiatric disease, saying that there truly are discrete disorders, we just tend to group them broadly and arbitrarily. Similar arguments can be construed for the structure of psychological traits like personality and intelligence.

In addition, there are often unresolved questions about the genetic architecture implied by the available evidence. For example, researchers used to believe that selection would reduce genetic variation in fitness traits, driving associated variants to fixation. This seemed to be borne out by low heritability coefficients. However, when researchers realized that fitness traits present a large target for mutation (Merilä & Sheldon, 1999), they reexamined the same heritability data expressed as the mean-standardized coefficient of variation (an absolute measure) and obtained large estimates of genetic variation. Heritability expressed as a proportion of total variation (a relative measure) had only appeared small in comparison, dwarfed by the large environmental variation (Miller & Penke, 2007). The conceptualization of fitness traits effectively reversed through a more appropriate statistic for variation.

Our understanding of how the forces of evolution shape traits' genetic architectures will continue to evolve. Thus, we begin with mechanisms potentially maintaining genetic variation before we discuss methods to identify causative genetic variants.

Forces of Evolution

We begin by introducing four basic forces that affect genetic variation in populations.

Mutation

All existing genetic variants once arose by mutation. Relative to the 6.4 billion base pairs of the human genomic sequence, mutations are rare events. Beneficial mutations are the rarest of all, the majority likely being neutral to fitness, with deleterious mutations making up the rest. Because the idea of a neutral mutation can be reduced to chance (or *drift*) being more important for its fate than selection, calling a mutation neutral also depends on its commonness, not just its effect size. A mutation with a small beneficial effect will have its fate determined mostly by chance while it is rare, because chance events can eliminate all copies. Once its frequency rises and in larger populations drift becomes relatively less important, so the mutation will be governed more by selection (Lanfear, Kokko, & Eyre-Walker, 2014).

The most common mutational event in humans is the change of a single base pair (the letters of the DNA), but there are also deletions, duplications, and insertions of base pairs or even longer parts of DNA (copy number variants). Aneuploidies (chromosomal aberrations), such as the duplication of chromosome 21, which causes Down syndrome, are rare but massive, accounting for most altered base pairs per birth. Except for aneuploidies, which are well known to exponentially increase in frequency with advancing maternal age, all types of mutations occur more often on the paternal side, and increasingly so with advancing paternal age at conception (Campbell & Eichler, 2013). Proximately, this is often attributed to the continuous division of cells in the paternal but not maternal germline (Kong et al., 2012), but ultimate explanations such as Bateman's principle (male investment in each offspring is lower) should be kept in mind (Stearns, 2005).

Selection

Selection occurs when there is heritable variation in fitness. Natural selection is frequently broken down into different subcategories. One grouping distinguishes positive, directional selection (favoring increases), disruptive selection, (favoring extremes), and stabilizing selection, (favoring decreased variation in a trait). Another grouping considers survival and sexual selection separately. Sometimes this is differentiated further into "episodes of selection." Survival selection could, for example, be divided into the chances of an ovum to be released in ovulation, sperm fertilizing an ovum, a zygote implanting, the pregnancy being carried to term (Stearns, 2005), surviving birth, living to

reproductive age, and further. Sexual selection might be divided into the odds of finding and attracting a mate, outcompeting same-sex rivals, the number of mates, the number of offspring per mate, and the fitness and number of offspring in the next few generations. Often the mistaken impression that selection has diminished in humans is, on closer inspection, limited to factors affecting perinatal and postnatal survival selection, with little heed paid to components of sexual selection.

Correlated Selection, Genetic Hitchhiking, and Pleiotropy

Genetic variants are not independently selected for. As the term "genetic hitchhiking" vividly implies, alleles can hitch a ride on the coattails, or *haplotype*, of a neighboring allele that is being selected for or against. The chances of inheriting a specific gene from a parent are not independent from those of its neighbors because we inherit genes in chunks. Over generations, recombination breaks haplotypes apart. Long, unbroken haplotypes signal strong recent selection for a new mutation, because the neighboring alleles of a beneficial mutation are "swept" along on the coattails before recombination can break them apart (known as a "hard sweep"). Shorter unbroken haplotypes can signal selection on standing (preexisting) genetic variation ("soft sweeps"; Pritchard, Pickrell, & Coop, 2010). Two or more alleles that usually co-occur (are in "linkage disequilibrium") and thus form a haplotype can have different, even opposing effects on fitness. Until recombination breaks them apart, they cannot be selected for independently.

Alleles experience correlated selection not only through proximity. Even a variant at a single locus can have multiple, *pleiotropic* effects on fitness via different phenotypic consequences. It can also make sense to distinguish fitness effects of an allele in different episodes of selection. For example, a mutation may be selected for pre-meiotically in the testes, but lead to Apert syndrome later on (Choi, Yoon, Calabrese, & Arnheim, 2008).

Genetic Drift

Luck plays a lead role when numbers are small. If there are few carriers of even a highly beneficial genetic variant, random events can eliminate all of them. Similarly, a deleterious variant can be fixated by chance, or a beneficial rare variant can randomly get lost in recombination. Either way, a gene variant may drift to fixation or extinction just by chance. If all variants at a locus are common (because no single variant is infrequent and the population is large), the law of large numbers implies that it will take long before either drifts to fixation. In humans, a comparatively extremely low genetic diversity points to genetic *bottlenecks* having been an important instance of drift (Gazave, Chang, Clark, & Keinan, 2013). Bottlenecks may occur through migration, such as when founder populations emigrated to North America, or when population sizes decreased dramatically through harsh conditions such as droughts,

epidemics, or ice ages. If the resulting population is small and not diverse (e.g., a clan), even beneficial alleles from the parent population may be lost through drift.

Gene Flow (or Migration)

When individuals carrying certain alleles move from one group to another, the frequency of alleles in each group also changes. This process is distinguished from unsystematic genetic drift, because relevant genetic variants may differentially influence the propensity to migrate and the success in each group and environment.

Maintenance Mechanisms

Prolonged directional or stabilizing selection on a trait will deplete its genetic variance. The mechanisms that maintain heritable variation in a trait can be understood as equilibria or trade-offs between the forces of evolution that change allele frequencies: selection, mutation, genetic drift, and gene flow. In some cases, it may seem as if evolution should lead to alternative genetic architectures with fewer trade-offs. Note that evolution is not over and that optimal solutions may not always be sufficiently better to be selected over merely adequate ones, which is, for example, why we still have blind spots in our eyes.

Mutation-Selection Balance (MSB)

Mutations continuously emerge. If they are entirely neutral, they are invisible to selection and may drift or hitchhike to extinction or fixation. But if they are deleterious, purifying selection will act against them. We rarely hear of dominant lethal mutations because they tend to be eliminated within one generation. Huntington's disease, which develops after the age of reproduction, is one example to the contrary.

If a trait is genetically complex, as most traits of interests to evolutionary psychologists likely are, many genes will be involved, not all of which play a crucial role. Hence, some deleterious mutations will be selected against less intensely and might linger for a few generations. If the mutational target size of a trait (the number of associated genetic loci) is large, mutations affecting the trait will accumulate, so that individuals carry a certain mutational load. Thus, variation in a trait such as physical attractiveness can be maintained even though it is likely under directional selection. In research on the genetics of autism spectrum disorders, new mutations appear to explain about 15% of cases (Devlin & Scherer, 2012), though this should not be equated with the part that MSB plays for autism, which may well be larger owing to older, inherited

mutation load. Debate revolves around the number of genes likely to be involved in a trait and on the question whether rare, recent or common, older mutations mostly disrupt such genes (Gazave et al., 2013).

Mutations in Balance With Stabilizing Versus Directional Selection

Traits under mutation-selection balance can be meaningfully differentiated further. If increases in a trait are linked to increased fitness (directional selection), new mutations should usually cause a decline in the trait. This assumption is implicit in most studies of MSB.

If fitness is instead linked to a certain optimum in a trait, it is said to be under stabilizing selection. Stabilizing selection acts to increase robustness to deleterious mutations, for example, by increasing genetic redundancy. For sexually recombining species, such as ours, it has also been suggested that increased mutational robustness need not imply a decrease in the evolvability of a trait (its potential to react to selection): Redundancy reduces the selective pressure on individual variants and thus allows variation to build up in the backup copy, creating a playground for genetic innovation. In this case, new mutations should cause comparatively smaller deviations from the optimum and might lead us to miss genetic associations if we focus on directional declines. The optimum would be expected to be the mean of a trait, at least in traits that were not subject to recent environmental changes. The shape of the eye might be an example of this exception: Myopia (shortsightedness; elongated eyes) is more common than hyperopia (early-onset farsightedness; shortened eyes), but the preponderance of myopia sufferers might be attributed to changes in our environment, in which near work became common and time outdoors decreased (Mingroni, 2004). To determine the not immediately visible optima of psychological traits, researchers could draw on associations of trait levels with survival and mate preferences as proxies of fitness consequences.

Balancing Selection

We now introduce a class of balancing mechanisms. In all of them, one selective pressure is counteracted by another in a different location, time, developmental stage, social environment, or intraindividual genetic context.

By Spatial Environmental Heterogeneity (Migration-Selection Balance)

Humans can experience different selective pressures in different environments. Selection by location need not be limited to selective pressures such as varying solar intensity (Norton et al., 2007) or altitude (Simonson et al., 2010), though these examples are best characterized.

Because personality may affect one's penchant for travel, migration can support spatial balancing selection: If those who want to see the world keep leaving their home island for the mainland, the remaining islanders may end up less open to experience on average (Ciani & Capiluppi, 2011). Selection would also reduce variance in openness if sedentary islanders did not occasionally interbreed with visitors from the mainland. This sort of recurring gene flow can maintain variation in openness. Similarly, sociability supports migration tendencies from rural to urban areas (Jokela, Elovainio, Kivimäki, & Keltikangas-Järvinen, 2008). In scenarios such as these, genetic variation is maintained because people within a population select themselves into the environments for which they are best adapted. Such niche picking (also known as active gene-environment correlation) is potentially a strong force in the maintenance of genetic variation in humans (Penke, 2010). In the population as a whole, no trait or underlying genetic variant would effectively be favored; thus, the selective pressures would balance.

Because cultural and other environmental explanations are hard to disentangle from genetically based psychological differences between populations, we advocate a cautious approach to this controversial topic. Some jump to premature conclusions about major genetic differences and even superiority based on flimsy evidence such as fairly high within-group heritability coefficients, but a balanced view of the evidence shows how difficult explaining group differences genetically is (Berg & Coop, 2014).

Because of humans' ecological dominance and concomitant capacity to shape the environment to their needs (niche construction), Penke et al. (2007; Penke, 2010) argued that the most important fluctuating aspect that humans need to adapt to is their social environment.

By Social Environment (Negative Frequency-Dependent Selection)

There are three *morphs* (types) of male common side-blotched lizards (*Uta stansburiana*), and three alleles at one Mendelian locus govern their throat color and concomitant behavior. Blue-throated males guard one mate and territory. Their mates can be stolen by larger, aggressive, orange-throated males, who keep large territories and multiple mates. Because they do not guard their mates well, they are vulnerable to having their mates stolen by yellow-throated males, who pretend to be female to sneakily gain access. This nontransitive mating game has been compared to rock-paper-scissors (Sinervo & Lively, 1996) and leads to oscillations in which the least common morph becomes more common in the next generation.

Biological sex is probably the most familiar morph under such negative frequency-dependent selection (NFDS) in humans, as the rarer sex becomes more desirable and thus has reproductive advantages due to mating market forces (Del Giudice, 2012). NFDS has also been invoked to explain primary psychopathy (Mealey, 1995), personality traits (Penke et al., 2007), and,

perhaps most fruitfully, immunity to parasites (Sutton, Nakagawa, Robertson, & Jamieson, 2011).

If psychopathy were under frequency-dependent selection, we might, through altered policy, lower the equilibrium frequency of psychopaths within few generations (Mealey, 1995).

Over Time (Generations)

If selection fluctuates over time more quickly than is needed for trait alleles to be driven to either fixation or extinction, variation can be maintained in oscillations. For example, if sex ratios in populations naturally fluctuate over time, genetic variation in personality traits that lead to better mating outcomes in one sex can be maintained by balancing selection (Del Giudice, 2012). If the fluctuations are predictable, selection should act to create genetically fixed conditional (facultative) strategies instead, a rich topic for life history theory (Nettle, Frankenhuis, & Rickard, 2013; Penke, 2009, 2010).

Over Time (Ontogenetic Development)

Earlier, we mentioned an allele that proliferates in the testes but leads to disease (Choi et al., 2008). Negatively correlated selection across developmental stages is also plausible for quantitative traits. For instance, large heads may support cognitive ability in later life, but they complicate birth (Miller & Penke, 2007). Selection should favor traits that are not subject to such trade-offs, but especially in conjunction with fluctuations of the fitness effects at different developmental stages, variation could be maintained.

By Genetic Variant at Other Loci (Epistasis)

An allele may have a beneficial or deleterious effect only in the presence or absence of other genetic variants. The sheer complexity of considering all the interactions in conjunction with the already large number of variants in the human genome has led some to propose that evolution would lead to mainly additive and even modularized variation in certain traits (W. G. Hill, Goddard, & Visscher, 2008), but epistasis might also be missed owing to insufficient statistical power.

By Genetic Variant at the Same Locus (Overdominance, Heterozygote Advantage, Selection-Drift Balance)

Consider a polymorphism, such as the one involved in sickle-cell anemia. Two copies of the polymorphism make blood cells sickle-shaped under low-oxygen conditions and typically lead to premature death. But having only one copy (heterozygosity) confers greater resistance to malaria. Individuals from areas in which malaria was a strong selective pressure are more often

carriers of the sickle-cell polymorphism. Heterozygotes have a selective advantage over homozygotes with either allele and so the sickle-cell allele can persist in the population at equilibrium frequency.
These equilibria are not stable: An allele that has the benefits but not the disadvantages will easily displace its competitor. We expect to see overdominance especially under strong, recent selection, such as that incurred by epidemics.

Mutation-Drift Balance (Selective/Ancestral Neutrality)

If mutations affecting a neutral trait arise so frequently that some linger before they drift out of existence, we expect genetic variation in this trait to linger as well. Because of the nature of genetic drift, existing, entirely neutral polymorphisms would linger longer in large populations. Because most human DNA is nonfunctional junk, which is not conserved through purifying selection, most mutations are neutral (Graur et al., 2013). One's first intuition might then be that most human individual differences are selectively neutral or "evolutionary noise" (Tooby & Cosmides, 1990). However, a commonly variable trait that is phenotypically visible to selection is less likely to be entirely neutral. This is especially the case since we tend to be interested in traits *because* they have predictive value for consequential life outcomes such as reproductive success, and thus evolutionary fitness. Additionally, because populations are larger nowadays, selection is more efficient, and will more often be stronger than drift (Penke et al., 2007).

In humans, with their rapidly changing culture and environment and with their rapidly increasing population size (Gazave et al., 2013), we might want to pay special heed to traits that used to be selectively neutral or nearly so, but no longer are. These are traits where we might expect natural selection to rapidly deplete genetic variation. Because traits under mutation-drift balance have a repository of standing variation and because selection is stronger than drift, it can decrease previously maintained variation.

A potential candidate for an ancestrally neutral psychological trait may be our preference for rising early or late: Our circadian rhythm is entrained to a universal source of light, the sun, in areas with little artificial light, where little time is spent indoors. With more artificial light, individuals' circadian rhythms become more variable (Wright et al., 2013), and such differences are moderately heritable (Barclay, Eley, Buysse, Archer, & Gregory, 2010). Possibly what we see here is cryptic genetic variation, revealed only under artificial light. Without it, the lack of variation in light exposure within populations might have meant that heritable differences were not visible, even though psychological differences that would have influenced self-exposure to artificial light already existed.

Mechanisms Implicating More Than One Trait a Time

In this section, we consider mechanisms that lead to the impression that there is heritable variability in a trait, but which are best understood in conjunction with other mechanisms and traits.

Mechanisms Related to Pleiotropy and Hitchhiking

When genes are pleiotropic (affect multiple traits) or in linkage (in close proximity to each other on a chromosome), genetic correlations among traits can appear. There are ways to discover genetic correlations and to analyze contemporary selection on multiple correlated traits (Stearns, Byars, Govindaraju, & Ewbank, 2010), but few studies have tried to do so for human evolutionary history.

The best-characterized examples of antagonistic pleiotropy arise in conjunction with biological sex. Traits like facial masculinity may be more adaptive in one sex than the other, but the respective alleles spend half their careers in each sex (A. J. Lee et al., 2014). Another important class of pleiotropic interactions may arise through the body's limited energy budget, especially that available for immune, brain, and gut functions. As a consequence, selection cannot optimize either trait, eventually resulting in a continuum of equally fit trait combinations maintained in the population.

Reactive Heritability

Not every trait with heritable individual differences needs to be subject to some sort of balancing mechanism itself. Instead, it could be calibrated to another heritable trait (Tooby & Cosmides, 1990). For example, Lukaszewski and Roney (2011) posited that extraversion might be calibrated to one's physical attractiveness and strength. Hence, we would find the signature of mutation-selection balance when studying extraversion in isolation, but would come to different conclusions when examining developmental and situational calibration of extraverted behavior to one's relative strength and attractiveness.

If they are not fixed at birth, we should not presume the primacy of physical traits. For example, we know that myopia appears to be linked to the amount of time children spend outdoors (Sherwin et al., 2012), but the substantial heritability estimates for myopia have led some researchers to downplay environmental explanations for the recent increase in myopia incidence (Mingroni, 2004). But if myopia heritability is partly reactive to children's heritable proclivity for outdoor play and if some children spend less time outdoors in recent times, which is plausible, these findings could be reconciled.

The Evolutionary Genetics Toolkit

In this section, we introduce the growing toolkit that is available to evolutionary geneticists. These tools were assembled from both quantitative and molecular genetics, as well as evolutionary psychology. We note what these tools can be used for, and how they are sometimes misused, but acknowledge how all of these methods make their contributions.

Twin and Family Studies

Twin studies are one of the oldest tools available and have withstood the test of time (Conley, Rauscher, Dawes, Magnusson, & Siegal, 2013). They rely on the key difference between monozygotic (identical) twins and dizygotic (fraternal) twins: Identical twins share all of their genes, while fraternal twins share on average half of the genes that were variable between their parents. A central result from twin studies is usually a heritability estimate, though the rich data from twin and family studies can answer many other questions too. The concept hails from plant and animal breeding, where it is used to predict response to artificial selection.

Estimates of heritability derived from twin studies have held up remarkably well when reexamined using different family relationships (e.g., parents, siblings, half- and adopted siblings) and can be easily extended to novel data such as the sometimes numerous offspring of sperm donors. In cases where selection is fairly clear-cut, estimates of heritability have borne out their usefulness as predictors of the response to selection. For example, children of sperm donors are taller in a manner consistent with their mothers' selection on donor height (J. C. Lee, 2013).

Usually things are not so tidy: Heritability estimates from twin studies often include some nonadditive variation, that is, variation that will not "breed true" to the next generation. Moreover, environmental confounds can make it hard to isolate an effect of selection, as the initiators of the Scottish Mental Survey discovered in 1947 when they attempted to show a decline of intelligence through differential fertility and found an increase instead (Ramsden, 2007). Humans simply do not behave like crops on a field or cattle in a breeding facility; they actively choose mates and both choose and modify their environments. This decreases the value of heritability estimates as more than a proof that genetic differences play a role in observable phenotypic variation (Johnson, Penke, & Spinath, 2011).

High heritability in twin studies has often been misunderstood to imply that a trait cannot be changed. To the contrary, species-typical universals such as two-leggedness have virtually zero heritability, because the underlying genes rarely vary. On the other hand, some gene-environment interactions were not apparent before the relevant environment changed: For example,

developing phenylketonuria, a disease causing intellectual disability, depends on consuming phenylalanine, which was a universal part of our diet before its damaging effects in some individuals became known.

Linkage Studies

Linkage studies, which identify larger genetic segments that segregate according to disease status in a pedigree, have been useful tools in the identification of "simple" Mendelian disorders, where single genes have major effects. They might also help once we learn to tell apart phenotypically similar diseases that we now group as complex psychiatric disorders (Mitchell, 2012). Linkage studies for most psychological variation have been characterized as a let-down. Still, they ruled out a suggested genetic architecture: If there were, for example, a single genetic locus causing human psychopathy (i.e., an exploitative social strategy) in analogy with the aforementioned sneaky side-blotched lizard, linkage patterns would have led to its identification.

Candidate Gene Studies

Candidate gene studies look for the association of a specific genetic locus with the trait of interest. By hypothesizing which locus may be involved *a priori*, they avoid correcting for multiple comparisons and can thus use smaller samples than the similar, but exploratory genome-wide association paradigm. They have come under intense criticism because of nonreplications and general doubts whether there is sufficient theory to predict candidate genes (Ioannidis, Trikalinos, Ntzani, & Contopoulos-Ioannidis, 2003).

Some recent studies, however, successfully employ candidate gene approaches, implicating candidate gene sets and apparently building on stronger theory than before. For example, W. D. Hill et al. (2014) reported and replicated an association of intelligence with variation in genes involved in one of the postsynaptic density complexes that have been implicated in cognitive functioning. Through preregistration of candidate genes, researchers could easily end disagreements and distrust whether their studies deserve the label of confirmatory research and concomitant relaxation of false discovery rates. Unfortunately, this is seldom done.

Genome-Wide Association Studies (GWAS)

GWAS assess the status of individuals on around a million genetic loci across the genome that are commonly variable in the population. While GWAS directly assess only around 0.033% of the human genome this way, linkage disequilibrium makes the assessed variants fairly exhaustive markers of common genetic variation, which is then related to the variation in the trait of interest. GWAS require large samples and have been early adopters of harsh

significance thresholds to account for the number of multiple comparisons (Ioannidis et al., 2003).

GWAS have been successful in the identification of some of the genes that matter for pigmentation, some medical disorders, height, and recently, schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Yet, for most psychological traits, especially normal variation, they rarely identified replicable associations (Chabris et al., 2012, 2013). This is often framed negatively, but GWAS effectively ruled out genetic architectures involving few common variants of medium-to-large effects for all psychological traits studied this way so far. Some researchers have advocated ever larger samples in order to potentially identify huge sets of genetic variants with individually miniscule effect sizes, while others argue that theory predicts only effects of questionable practical relevance and that family-based designs are better suited (Mitchell, 2012).

Using Sequenced Exomes and Genomes in Association Studies

Sequencing refers to identifying every single base pair in someone's genome, not just a few commonly polymorphic loci, as in GWAS. When sequencing is limited to protein-coding genes (ca. 1%–2% of the whole genome), this subset is called the exome. The exome constitutes a more manageable amount of data and has been considered promising for clinical variation. However, much of it is conserved between species and a lot of recent selection has operated on promoters outside the exome (Enard et al., 2014; The 1000 Genomes Project Consortium, 2012), making exome variation a less likely candidate for contributing to the genetic architecture of psychological traits in the normal range (Marioni et al., 2014).

With the amounts of data generated by genome sequencing, entirely exploratory research would not be useful due to the sample sizes required to filter chance findings. Integrating prior knowledge, such as annotations on regions with a signature of recent selection or expression in the brain (Ma et al., 2013), or alternatively relying on summary indices of rare genetic variants, a direct operationalization of mutation load (Marioni et al., 2014), may make such data manageable.

Genomic Prediction and Genome-Wide Complex Trait Analysis (GCTA)

A method formerly used primarily to predict breeding value in domestic animals has recently become popular in human genetics under the name GCTA (Yang et al., 2011). The general method estimates distant relatedness (less than fourth cousins) between individuals in the general population on the basis of common genetic variants, as provided by GWAS. Unlike GWAS, this method

does not identify individual important loci. Instead, the distant relatedness is used to infer a heritability score akin to that known from twin studies, but based solely on molecular data. After many GWAS failed to identify loci associated with psychological traits, GCTA provided a means of showing that the genotype data was actually informative: It can validate heritability estimates and be used to enable marker-assisted breeding (though this application is unlikely in humans), even if it does not identify causative genes and hence provides no foothold to find biological pathways. A frequently raised objection is that GCTA heritability estimates might be spurious, driven by the resemblance of distantly genetically related individuals for nongenetic reasons, such as similar environments because of shared ancestry and migration history. Researchers working with GCTA acknowledge such confounds, and the discussion revolves mostly about whether the corrections are sufficient (Conley et al., 2014; Yang et al., 2011).

Some researchers also doubt whether finding high GCTA heritability implies that the infinitesimal model of many common variants of tiny effect applies, especially when debilitating disorders are under study (Mitchell, 2012). Maybe more agreement can be fostered by a shift to delineating a fully featured genetic architecture, acknowledging the balanced forces enumerated in this chapter.

Paternal Age Effects

By sequencing and comparing the genomes of both parents and an offspring, Kong et al. (2012) convincingly demonstrated that the number of newly occurred single nucleotide variants in offspring can almost entirely be accounted for by the father's age at conception. Thus, paternal age can be used as a proxy variable to infer the effect of new mutations. To isolate this effect, the fact that human reproductive timing is not governed by chance has to be statistically controlled. Initially reported negative associations between paternal age and intelligence in the normal range (Malaspina et al., 2005) have not been replicated in later studies. Controlling parental intelligence, an important predictor of reproductive timing, may account for some of the observed heterogeneity of effects (Arslan, Penke, Johnson, Iacono, & McGue, 2014). Employing sibling comparison designs also led to the disappearance of paternal age effects on intelligence, while a strong association with attention deficit hyperactivity disorder became visible only with sibling controls (D'Onofrio et al., 2014).

Properly isolated, paternal age effects can provide evidence for a trait being under mutation-selection balance. In addition, they can be useful to predict the effect of increasingly delayed reproduction in the industrialized world on average mutation load (Sartorius & Nieschlag, 2010).

Genome and Exome Triplets and Quads

When the entire exomes or even genomes of parent-offspring trios are sequenced, it becomes possible to count new mutations, that is, alleles that neither parent carried. By assessing which haplotype a mutation lies on, it is also possible to identify the parent of origin. Then, mutation counts can predict, for example, intellectual disability (Rauch et al., 2012) and recurring mutations can be used to zero in on causative genes.

Exome quads (both parents and two offspring) have been used in autism genetics. Using genome annotations, Iossifov et al. (2012) estimated which mutations interrupted genes. By also sequencing unaffected siblings whose genomes were recombined from a common parental pool, they could isolate the effect of having more disrupted genes. Studies on autism genetics tried to isolate the effect of new mutations from assortative mating by considering only families without a familial history of autism and through sibling comparisons. These molecular genetic studies corroborate earlier results of autism increasing with paternal age.

Inbreeding Depression and Outbreeding Elevation

Inbreeding depression refers to a fitness decrease in offspring of consanguinous unions. Consanguinous parents (second cousins and closer) and their offspring make up about 10% of the world's population, though their prevalence has been predicted to decline (Bittles & Black, 2009). Franssen (2009) reported a linear negative relationship between offspring mental ability and consanguinity ranging from second-cousin marriages to incest. Such associations are confounded by many unobserved common causes. For example, lower parental education can, via lower mobility, increase the likelihood of marrying relatives and thus inflate estimates of inbreeding depression. The family history and cultural prevalence of consanguinity (e.g., in clans and castes) affect inbreeding coefficients too, so that estimates based on just two generations can be off (Bittles, 2010).

Outbreeding elevation, also known as *hybrid vigor* or *heterosis*, refers to the increased phenotypic quality of the offspring of genetically more distant parents. This phenomenon is very familiar to plant and animal breeders. Mules may be the most iconic hybrids and hybrid maize the most frequently consumed. The *vigor* does not necessarily translate to evolutionary fitness: Mules are valued beasts of burden but are frequently infertile. This is because too-distant genetic relationships between parents can break up co-adapted gene complexes during recombination, hence breaking vital functions such as the ability to reproduce. A bit of both may have happened when modern humans and Neanderthals interbred (Sankararaman et al., 2014). Hybrid vigor can also occur when inbreeding ends: Mixed-breed dogs have higher life expectancy

than most purebreds (O'Neill, Church, McGreevy, Thomson, & Brodbelt, 2013). Mingroni (2004) proposed that urbanization and generally less sedentism led to decreased inbreeding and might be partial causes for the recent increases in height and intelligence in industrialized countries.

Runs of Homozygosity

Analogously to GCTA, which employs DNA-based subtle relatedness to validate twin studies' estimate of heritability, runs of homozygosity (ROH) are an attempt to characterize subtle inbreeding on a molecular level. If long stretches of a diploid genome are homozygous, that is, both strands of DNA have the same variants, we can infer that closely related individuals have bred. If many shorter stretches are homozygous, we can infer ancient relatedness (Kirin et al., 2010). The genomic approach has the benefit that inbreeding over several generations can be characterized, though it is important to supplement this with knowledge of the history of endogamous marriage, founder effects, and population bottlenecks (Bittles, 2010). Homozygosity appears to play a role not only in well-characterized recessive disorders such as cystic fibrosis, but also for traits like personality (Verweij et al., 2012, 2014). Power et al. (2013) found a zero-to-slightly-positive association between ROH burden and intelligence, which conflicts with (possibly more biased) pedigree-based estimates of inbreeding effects (Franssen, 2009).

Relations With Fitness (Lifetime Reproductive Success) and Mate Preferences

It may seem as if we have so far neglected the obviously relevant effects of traits on fitness measures in this chapter. This is because, with some exceptions (e.g., pervasive developmental disorders), it is difficult to establish that the same association has persisted over evolutionary time and is thus indicative of the balancing mechanism that primarily upheld variation in a trait. We lack historical data for psychological traits, and many associations between normal variation and fitness estimated nowadays could be fickle. Contemporary selection on human individual differences is interesting in itself (Stearns et al., 2010), but we expect evolutionary genetics, among other disciplines, to answer the question "Why did humans evolve to be this way?"

In the age of widespread, effective contraception, it can be argued that mate preferences and choices are better-preserved indicators of sexual selection than correlations with reproductive success. In addition to being more immediately assessable than lifetime reproductive success, mate preferences have been shown to be relatively culturally invariant (Buss, 1989), unlike total fertility. Perinatal and postnatal survival selection plausibly have decreased in intensity since the advent of hygiene, modern health care, less frequent infanticide, and lower infant and maternal mortality. Still, a large number of

pregnancies are not carried to term and many debilitating, previously lethal genetic conditions, such as severe disability, may now be sexually selected against owing to lower attractiveness in the mating market.

Correlations With Indicators of Developmental Stability

Bilateral fluctuating asymmetry (FA) of the body is presumed to be an indicator of developmental stability, operating under the assumption that mutation-free organisms in good condition will be more symmetrical (Polak, 2003). Correlations with FA are thus assumed to provide an indirect way to tap a trait's association with mutation load. This paradigm is prevalent in evolutionary psychology and somewhat plagued by publication bias (Van Dongen & Gangestad, 2011). Hardly any studies take a molecular or population genetic approach to fluctuating asymmetry in humans. Future studies should more directly examine an association of developmental stability indicators with rare genetic variant burden, paternal age, or consanguinity before correlations with FA can be deemed valid proxies for tapping "good genes." Preregistration of studies could foster greater trust, especially that of scientists in adjacent domains such as genetics.

Conclusion and Outlook

Evolution by natural selection occurs as long as there is heritable variation related to differential fitness in the population. The evidence for both is ubiquitous even today, posing the question why so much genetic variation persisted. Genetic variance is influenced by mutation, selection, drift, and migration, and combinations of these four forces can yield balanced states in which it is maintained. This has been known since the modern synthesis in the 1930s, but our understanding of the molecular genetics underlying these processes has radically progressed. We are increasingly able to learn about the genetic architecture underlying psychological traits. Although the resulting picture will not be as simple as most researchers assumed even a few years ago, it can eventually provide insights about the evolutionary history and the selective pressures currently acting on these traits (Penke et al., 2007).

The evolutionary genetic tool kit includes complementary tools from molecular, behavior genetics and classical evolutionary psychology. Every available method has so many caveats that only converging evidence can enable us to single out theories as tenable. Unfortunately, even closely neighboring disciplines do not often lend each other tools and insights. For example, pure life history models of psychopathology (Del Giudice, Klimczuk, Traficonte, & Maestripieri, 2014) are inconsistent with the accumulating evidence that mutation load plays a major role in the autism and schizophrenia

spectra (Andreassen et al., 2014). Research on runs of homozygosity and mutation load could verify assumptions inherent in studies on fluctuating asymmetry. We need to subject our favored evolutionary explanations to tools from outside our own respective fields. Different disciplines can find it hard to properly evaluate and trust results outside their own field, especially if there is publication bias. Data and discussion brought to bear on the matter may have ideological baggage and bias (Ramsden, 2007), as researchers on, for example, intelligence or inbreeding, where science is easily conflated with moral judgments, know well. However, we can restore trust in areas plagued by bias (e.g., candidate gene and fluctuating asymmetry studies) through preregistration, replication, collaboration in consortia, and greater transparency. Such quality badges can be recognized even if the exact details are beyond us (Miller, 2011). By embracing such superior scientific standards we can protect our theories from the charge of being "just-so stories."

AU: Please add the Andreassen et al. source to the references.

RCA: Done.

It is encouraging, however, that all these approaches share a common evolutionary meta-theory, which could help to integrate knowledge acquired using diverse tools and build a common understanding. We have referenced numerous positive examples throughout this chapter. Mutual assistance and understanding should lead not only to agreement on the existence of heritable individual differences, but on the *mechanisms* maintaining them. Even where we identify genetic architectures that make it hard for us to detect important causative genes (e.g., an infinitesimal number of causative genes of small effect, genetic heterogeneity, or epistasis), there is a lot to be gleaned from understanding maintaining mechanisms. These mechanisms are not idle theory; they have practical applications. Policy and mores already exert influence on demography, reproductive timing, and selective pressures. We do not need to know specific genetic variants to predict what will happen to autism incidence if people reproduce later, nor to characterize the role of assortative mating and consanguinity in the age of online dating, nor to understand the impact of anciently constant selective forces suddenly swayed by new technology.

Where we identify traits with a genetic architecture conducive to identifying causative genes, many doors open for vertical integration (Y. W. Lee, Gould, & Stinchcombe, 2014) with biology and neuroscience: We can study pathways, develop drugs and genetic screenings, examine molecular signatures of selection and demographic history (Enard et al., 2013), use Mendelian randomization techniques (Smith & Ebrahim, 2004) to identify modifiable causes of disease, and make inferences about earlier hominids' psychological characteristics on the basis of shared polymorphisms.

Darwin knew nothing about the genetics underlying evolution, but our ever more detailed understanding allows us to fully embrace the potential of merging evolutionary theory with genetics. Evolutionary genetics enriches evolutionary psychology by providing a theoretical framework and tools to

integrate individual differences and recent evolution (Penke, 2010), and thus ultimately an understanding of why we are the way we are and how we became that way.

References

Andreassen, O. A., Harbo, H. F., Wang, Y., Thompson, W. K., Schork, A. J., Mattingsdal, M., … Dale, A. M. (2014). Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Molecular Psychiatry, 1-8.* Advance online publication.

Arslan, R. C., Penke, L., Johnson, W., Iacono, W. G., & McGue, M. (2014). The effect of paternal age on offspring intelligence and personality when controlling for paternal trait level. *PLoS ONE*, *9*, e90097.

Barclay, N. L., Eley, T. C., Buysse, D. J., Archer, S. N., & Gregory, A. M. (2010). Diurnal preference and sleep quality: Same genes? A study of young adult twins. *Chronobiology International*, *27*, 278–296.

Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, *10*, e1004412.

Bittles, A. H. (2010). Consanguinity, genetic drift, and genetic diseases in populations with reduced numbers of founders. In M. R. Speicher, A. G. Motulsky, & S. E. Antonarakis (Eds.), *Vogel and Motulsky's human genetics* (pp. 507–528). Berlin, Germany: Springer.

Bittles, A. H., & Black, M. L. (2009). Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences*, *107*, 1779–1786.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49.

Buss, D. M., & Hawley, P. H. (2011). *The evolution of personality and individual differences*. New York, NY: Oxford University Press.

Buss, D. M., & Penke, L. (2014). *Evolutionary personality psychology.* In L. Cooper & R. Larsen (Eds.), Handbook of Personality Processes and Individual Differences. Washington, DC: American Psychological Association.

Campbell, C. D., & Eichler, E. E. (2013). Properties and rates of germline mutations in humans. *Trends in Genetics*, 1–10.

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., … Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science*, *23*, 1314–1323.

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J. P., Glaeser, E. L., Borst, G., … Laibson, D. I. (2013). Why it is hard to find genes associated with social science traits: Theoretical and empirical considerations. *American Journal of Public Health*, *103*, S152–S166.

Choi, S.-K., Yoon, S.-R., Calabrese, P., & Arnheim, N. (2008). A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common human disease mutations. *Proceedings of the National Academy of Sciences, USA*, *105*, 10143–10148.

Ciani, A. C., & Capiluppi, C. (2011). Gene flow by selective emigration as a possible cause for personality differences between small islands and mainland populations. *European Journal of Personality*, *25*, 53–64.

Conley, D., Rauscher, E., Dawes, C., Magnusson, P. K. E., & Siegal, M. L. (2013). Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behavior Genetics*, *43*, 415–426.

Conley, D., Siegal, M. L., Domingue, B. W., Mullan Harris, K., McQueen, M. B., & Boardman, J. D. (2014). Testing the key assumption of heritability estimates based on genome-wide genetic relatedness. *Journal of Human Genetics*, *59*, 342–345.

D'Onofrio, B. M., Rickert, M. E., Frans, E., Kuja-Halkola, R., Almqvist, C., Sjölander, A., … Lichtenstein, P. (2014). Paternal age at childbearing and offspring psychiatric and academic morbidity. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2013.4525

Del Giudice, M. (2012). Sex ratio dynamics and fluctuating selection on personality. *Journal of Theoretical Biology*, *297*, 48–60.

Del Giudice, M., Klimczuk, A. C. E., Traficonte, D. M., & Maestripieri, D. (2014). Autistic-like and schizotypal traits in a life history perspective: Diametrical associations with impulsivity, sensation seeking, and sociosexual behavior. *Evolution and Human Behavior*, *35*, 415–424.

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*, 1285–1302.

Devlin, B., & Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Current Opinion in Genetics & Development*, *22*, 229–237.

Enard, D., Messer, P. W., & Petrov, D. A. (2014). Genome-wide signals of positive selection in human evolution. *Genome Research*, *24*, 885–895.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*, 1–25.

Franssen, D. (2009). *Is there a linear relationship between inbreeding and mental ability?: A meta-analysis*. Retrieved from http://dare.uva.nl/document/152307

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, *23*, 573–587.

Gangestad, S. W., & Yeo, R. A. (1997). Behavioral genetic variation, adaptation and maladaptation: An evolutionary perspective. *Trends in Cognitive Sciences*, *1*, 103–108.

Gazave, E., Chang, D., Clark, A. G., & Keinan, A. (2013). Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics*, *195*, 969–978.

Graur, D., Zheng, Y., Price, N., Azevedo, R. B. R., Zufall, R. A., & Elhaik, E. (2013). On the immortality of television sets: "Function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biology and Evolution*, *5*, 578–590.

Hill, W. D., Davies, G., van de Lagemaat, L. N., Christoforou, A., Marioni, R. E., Fernandes, C. P. D., … Deary, I. J. (2014). Human cognitive ability is influenced by genetic

variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Translational Psychiatry*, *4*, e341.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, *4*, e1000008.

Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., & Contopoulos-Ioannidis, D. G. (2003). Genetic associations in large versus small studies: An empirical assessment. *The Lancet*, *361*, 567–571.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., … Wigler, M. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, *74*, 285–299.

Johnson, W., Penke, L., & Spinath, F. M. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of Personality*, *25*, 254–266.

Jokela, M., Elovainio, M., Kivimäki, M., & Keltikangas-Järvinen, L. (2008). Temperament and migration patterns in Finland. *Psychological Science*, *19*, 831–837.

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*, *5*, e13996.

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, *488*, 471–475.

Lanfear, R., Kokko, H., & Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution*, *29*, 33–41.

Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2014). Genetic factors that increase male facial masculinity decrease facial attractiveness of female relatives. *Psychological Science*, *25*, 476–484.

Lee, J. C. (2013). *Quantitative genetics in the postmodern family of the donor sibling registry*. Retrieved from https://www.donorsiblingregistry.com/sites/default/files/files/dissertation(1).pdf

Lee, Y. W., Gould, B. A., & Stinchcombe, J. R. (2014). Identifying the genes underlying quantitative traits: A rationale for the QTN programme. *AoB Plants*, plu004.

Lukaszewski, A. W., & Roney, J. R. (2011). The origins of extraversion: Joint effects of facultative calibration and genetic polymorphism. *Personality & Social Psychology Bulletin*, *37*, 409–421.

Ma, L., Clark, A. G., & Keinan, A. (2013). Gene-based testing of interactions in association studies of quantitative traits. *PLoS Genet*, *9*, e1003321.

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, *35*, 303–339.

Malaspina, D., Reichenberg, A., Weiser, M., Fennig, S., Davidson, M., Harlap, S., … Knobler, H. Y. (2005). Paternal age and intelligence: Implications for age-related genomic changes in male germ cells. *Psychiatric Genetics*, *15*, 117–125.

Marioni, R. E., Penke, L., Davies, G., Huffman, J. E., Hayward, C., & Deary, I. J. (2014). The total burden of rare, non-synonymous exome genetic variants is not associated with childhood or late-life cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140117.

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, *18*, 523–541.

Merilä, J., & Sheldon, B. C. (1999). Genetic architecture of fitness and nonfitness traits: Empirical patterns and development of ideas. *Heredity*, *83*, 103–109.

Miller, G. F. (2011). Are pleiotropic mutations and Holocene selective sweeps the only evolutionary-genetic processes left for explaining heritable variation in human psychological traits? In D. M. Buss & P. H. Hawley (Eds.), *The evolution of personality and individual differences* (p. 376). New York, NY: Oxford University Press.

Miller, G. F., & Penke, L. (2007). The evolution of human intelligence and the coefficient of additive genetic variance in human brain size. *Intelligence*, *35*, 97–114.

Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence*, *32*, 65–83.

Mitchell, K. J. (2012). What is complex about complex disorders? *Genome Biology*, *13*, 237.

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20131343.

Norton, H. L., Kittles, R. A., Parra, E., McKeigue, P., Mao, X., Cheng, K., … Shriver, M. D. (2007). Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Molecular Biology and Evolution*, *24*, 710–722.

O'Neill, D. G., Church, D. B., McGreevy, P. D., Thomson, P. C., & Brodbelt, D. C. (2013). Longevity and mortality of owned dogs in England. *The Veterinary Journal*, *198*, 638–643.

Penke, L. (2009). Adaptive developmental plasticity might not contribute much to the adaptiveness of reproductive strategies. *Behavioral and Brain Sciences*, *32*, 38–39.

Penke, L. (2010). Bridging the gap between modern evolutionary psychology and the study of individual differences. In D. M. Buss & P. H. Hawley (Eds.), *The evolution of personality and individual differences* (pp. 243–279). New York, NY: Oxford University Press.

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality*, *21*, 549–587.

Polak, M. (2003). *Developmental instability: causes and Consequences*. New York, NY: Oxford University Press.

Power, R. A., Nagoshi, C., DeFries, J. C., Donnelly, P., Barroso, I., Blackwell, J. M., … Plomin, R. (2013). Genome-wide estimates of inbreeding in unrelated individuals and their association with cognitive ability. *European Journal of Human Genetics*. doi:10.1038/ejhg.2013.155

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, R208–R215.

Ramsden, E. (2007). A differential paradox: The controversy surrounding the Scottish mental surveys of intelligence and family size. *Journal of the History of the Behavioral Sciences*, *43*, 109–134.

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., … Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet*, *380*, 1674–1682.

Rholes, W. S., & Simpson, J. A. (2006). *Adult attachment: Theory, research, and clinical implications*. New York, NY: Guilford Press.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., … Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*, 354–357.

Sartorius, G. A., & Nieschlag, E. (2010). Paternal age and reproduction. *Human Reproduction Update*, *16*, 65–79.

Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, *20*, 99–103.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*, 421–427.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106*, 15073–15078.

Sherwin, J. C., Reacher, M. H., Keogh, R. H., Khawaja, A. P., Mackey, D. A., & Foster, P. J. (2012). The association between time spent outdoors and myopia in children and adolescents: A systematic review and meta-analysis. *Ophthalmology*, *119*, 2141–2151.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., … Ge, R. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science*, *329*, 72–75.

Sinervo, B., & Lively, C. M. (1996). The rock–paper–scissors game and the evolution of alternative male strategies. *Nature*, *380*, 240–243.

Smith, G. D., & Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology*, *33*, 30–42.

Stearns, S. C. (2005). Issues in evolutionary medicine. *American Journal of Human Biology*, *17*, 131–140.

Stearns, S. C., Byars, S. G., Govindaraju, D. R., & Ewbank, D. (2010). Measuring selection in contemporary human populations. *Nature Reviews Genetics*, *11*, 611–622.

Sutton, J. T., Nakagawa, S., Robertson, B. C., & Jamieson, I. G. (2011). Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology*, *20*, 4408–4420.

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*, 56–65.

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, *58*, 17–67.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*, 160–164.

Van Dongen, S., & Gangestad, S. W. (2011). Human fluctuating asymmetry in relation to health and quality: A meta-analysis. *Evolution and Human Behavior*, *32*, 380–398.

Verweij, K. J. H., Abdellaoui, A., Veijola, J., Sebert, S., Koiranen, M., Keller, M. C., … Zietsch, B. P. (2014). The association of genotype-based inbreeding coefficient with a range of physical and psychological human traits. *PLoS ONE*, *9*, e103102.

Verweij, K. J. H., Yang, J., Lahti, J., Veijola, J., Hintsanen, M., Pulkki-Råback, L., … Zietsch, B. P. (2012). Maintenance of genetic variation in human personality: Testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. *Evolution*, *66*, 3238–3251.

Wright, K. P., McHill, A. W., Birks, B. R., Griffin, B. R., Rusterholz, T., & Chinoy, E. D. (2013). Entrainment of the human circadian clock to the natural light-dark cycle. *Current Biology*, *23*, 1554–1558.

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., … Hayes, M. G. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, *43*, 519–525.

Appendix B.

Manuscript 2 (Older fathers' children have lower evolutionary fitness across four centuries and in four populations)

Title: Older fathers' children have lower evolutionary fitness across four centuries and in four populations

Forthcoming in: Proceedings of the Royal Society B: Biological Sciences

Authors: Ruben C. Arslan[1,2], Kai P. Willführ[3], Emma M. Frans[4,5], Karin J. H. Verweij[6,7], Paul-Christian Bürkner[8], Mikko Myrskylä[3, 9,10], Eckart Voland[11], Catarina Almqvist[5,12], Brendan P. Zietsch[7,13], Lars Penke[1,2]

Affiliations:

1. Biological Personality Psychology,
   Georg Elias Müller Institute of Psychology,
   Georg August University Göttingen,
   37073 Göttingen, Germany
2. Leibniz ScienceCampus Primate Cognition,
   37073 Göttingen, Germany
3. Max Planck Institute for Demographic Research,
   18057 Rostock, Germany,
4. Department of Psychiatry,
   University of Oxford,
   Warneford Hospital,
   Oxford, OX3 7JX, United Kingdom
5. Department of Medical Epidemiology and Biostatistics,
   Karolinska Institutet,
   171 77 Stockholm, Sweden

6. Department of Biological Psychology,
VU University,
1081 BT, Amsterdam, the Netherlands
7. School of Psychology,
University of Queensland,
St. Lucia, Brisbane, QLD 4072, Australia
8. Department of Psychology
University of Münster
48149 Münster, Germany
9. Department of Social Policy,
London School of Economics and Political Science,
London, WC2A 2AE, United Kingdom
10. Population Research Unit,
University of Helsinki,
00100 Helsinki, Finland
11. Department of Biophilosophy,
Justus Liebig University Gießen,
35390 Gießen, Germany
12. Pediatric Allergy and Pulmonology Unit at Astrid Lindgren Children's
Hospital, Karolinska University Hospital, Stockholm, Sweden
13. Genetic Epidemiology,
QIMR Berghofer Medical Research Institute,
Brisbane, QLD 4006, Australia

Corresponding author: Ruben C. Arslan

Biological Personality Psychology,

Georg Elias Müller Institute of Psychology,

Georg August University Göttingen,

Goßlerstr. 14,

37073 Göttingen, Germany

Tel.: +49 551 3920704

Email: ruben.arslan@gmail.com

Keywords: paternal age, evolutionary fitness, mutation, genetic load, reproductive success

Abstract

Higher paternal age at offspring conception increases *de novo* genetic mutations. Based on evolutionary genetic theory we predicted older fathers' children, all else equal, would be less likely to survive and reproduce, i.e. have lower fitness. In sibling control studies, we find support for negative paternal age effects on offspring survival and reproductive success across four large populations with an aggregate $N > 1.4$ million. Three populations were pre-industrial (1670-1850) Western populations and showed negative paternal age effects on infant survival and offspring reproductive success. In 20th-century Sweden, we found minuscule paternal age effects on survival, but found negative effects on reproductive success. Effects survived tests for key competing explanations, including maternal age and parental loss, but effects varied widely over different plausible model specifications and some competing explanations such as diminishing paternal investment and epigenetic mutations could not be tested. We can use our findings to aid in predicting the effect increasingly older parents in today's society will have on their children's survival and reproductive success. To the extent that we succeeded in isolating a mutation-driven effect of paternal age, our results can be understood to show that *de novo* mutations reduce offspring fitness across populations and time periods.

Media summary

Fathers' and mothers' average ages at birth are increasing throughout the developed world, though they are presently still on par with pre-industrial

reproductive timing. We find that children of older fathers have fewer children themselves in four populations across four centuries: three pre-industrial populations from the 17-19th century, and 20th-century Sweden (total sample size > 1.4m). A child gets most new genetic mutations from its father, which increase continuously with his age. We can use the father's age to indirectly learn about the effect of new mutations on the child, but some complicating factors could not be controlled.

Background

A child carries on average about 60 genetic *de novo* single nucleotide mutations (SNMs), which were not present in either of the biological parents' genomes [1,2]. Of those that are not functionally neutral, most reduce evolutionary fitness, as random changes to well-calibrated systems usually do [3,4]. Importantly, *de novo* mutations can be dominantly lethal or sterility-inducing early in life, unlike inherited deleterious variants. The older a father is, the more *de novo* mutations his child will tend to carry. This is dictated by the fundamental fact that cell replication engenders errors [5], and male spermatogonial, but not female oogonial stem cells, replicate frequently, beginning a regular schedule of one division per 16 days in puberty [6].

Kong et al. sequenced the genomes of parent-child triplets and quartets, so that they could pinpoint mutations and their parental origin [1]. They found that a child's number of *de novo* SNMs could be predicted very well (94% non-stochastic variance explained) by the father's age at the child's birth, henceforth *paternal age*. Mothers appears to transmit only a third to half as

many SNMs per year as fathers [4,7]. Thus, paternal age appears to be the main predictor of varying offspring *de novo* mutation load, in part because of its causal role and to a lesser extent because of its correlation with maternal age. SNMs are the most common mutational event, but copy number variants also increase with paternal age; other structural variants tend to come from the father too [8]. Aneuploidies (aberrant chromosome counts) are a well-known exception: they occur more often when older mothers conceive [2]. Subsequent studies have confirmed the central role of paternal age for mutations [4,6].

In clinical research, paternal age has shown usefulness as a placeholder variable for *de novo* mutations: after initial epidemiological studies reported paternal age effects on autism [9], sibling comparison studies confirmed they were not due to inherited dispositions [10]. Then, exome-sequencing studies corroborated the paternal age effects by directly counting mutations that were not present in either parent's exome and found a higher mutational burden in autistic children than in unaffected siblings [11]. These findings elucidated disease aetiology both from an evolutionary and a clinical standpoint, by explaining how an early-onset disease linked to very low reproductive success could linger in the face of natural selection.

Given the links enumerated above, paternal age should, via increased mutations, decrease offspring fitness. By fitness, we mean each offspring's average contribution to the gene pool of successive generations. We can

approximate this contribution through the offspring's number of descendants [12].

So far, most paternal age effect studies have focussed on medical, psychological and behavioural traits, such as physical and psychiatric disease, or intelligence [10,13–16]. Though many of these traits plausibly affect evolutionary fitness now, it is not always clear how they affected fitness before the 20th century. Moreover, there are scant records on such traits from this time, and they are not necessarily comparable to modern records. Births and deaths, or baptisms and burials, on the other hand, have been meticulously recorded in churches. Survival and reproductive success were and still are good measures of evolutionary fitness. Fitness is the most 'downstream' phenotype of all, in the sense that all non-neutral mutations affect it by definition [17].

Paternal age effects on mutations should in principle be universal across species, but nonhuman animal studies have thus far been restricted to birds [18,19] and have, with one exception [19], been studied under the broader topic of senescence, without attempts to separate mutational or epigenetic effects from behavioural effects of parental senescence on breeding capability. Studies on humans have examined isolated fitness components such as infant survival, longevity, marriage or reproduction in single populations in one place and at one time [20–23]. Some such studies have focussed on longevity, which has an ambiguous relationship to evolutionary fitness owing to life history trade-offs, such as trading off higher early-life

reproduction for earlier mortality [24]. Some have examined maternal age or birth order, but ignored paternal age [25]. Some focussed on environmental explanations, such as decreased parental investment [26], but these are not necessarily sufficient to explain paternal age effects. In wild house sparrows, the age of the biological parents had negative consequences even in a cross-fostering experiment [19]. Such experiments are not possible in humans, but we can statistically adjust for proxy measures of parental investment. In all, owing to variable methodology and sample sizes across studies, we cannot reliably compare findings to discover theoretically meaningful moderators.

The Present Study

Here we investigated paternal age effects on offspring fitness, focussing on the offspring's reproductive success, i.e. their number of children. To be able to compare all children of a father, we also included children who had no children themselves, even if they died young. Reproductive success is a good predictor of an individual's contribution to the next generation's gene pool [12]. In addition, we separately examined early survival, marriage success and reproductive success as successive episodes across the lifespan during which natural and sexual selection occur. Based on evolutionary genetic theory, we predicted that in aggregate we would find small, negative effects of paternal age on offspring fitness throughout the lifespan [27]. Some *de novo* mutations will have large negative effects early on, but many more will be (nearly) neutral. In aggregate, on the population level, this implies a small stochastically variable increase in deleterious effects with paternal age.

Because humans do not time their reproduction randomly, paternal age effects may be confounded by social and genetic factors [28–30] that are associated with both age of reproduction and offspring reproductive success. Because we aimed to isolate *mutation-driven* effects of paternal age as thoroughly as possible, we analysed the paternal age effect within full biological sibships and adjusted for a between-family effect. This effectively controls for many potential confounds. Full siblings share a parental gene pool, so that genetic load, which accumulated over generations, is distributed across them randomly. Siblings also usually share much of their early environment, and access to resources such as wealth and land. Because social convention may additionally link inheritance to birth order, we also adjusted for other social factors, such as birth order and parental loss. Additionally, we examined grandpaternal age effects where possible.

In doing so, we try to accomplish two goals: first, to isolate a potential biological, mutation-driven effect of paternal age on offspring fitness, and second, to compare different populations in different times and places, with high statistical power and comparable methodology.

Methods

Populations

To test our hypotheses before the turn of the 20th century, we used genealogies drawn from church records in the Saint-Lawrence valley, Québec (Canada), the Krummhörn (Germany) and four historical Swedish regions. To

compare these populations to 20th century Sweden, we used a population-based linkage study from Swedish national health registers. To ensure minimal censoring we drew subsets with adequately complete records.

We used computerized and linked registers of births (and baptisms), deaths (and burials) and marriages to reconstruct family pedigrees and life histories for individuals. We call the individuals whose father's age we compared with their siblings' "anchors" wherever it aids comprehension. Further descriptive statistics can be found in Table 1 and on the online supplementary website at https://rubenarslan.github.io/paternal_age_fitness/ [31].

	1720-1850 Krummhörn	**1670-1750 Québec**	**1760-1850 Sweden**	**20th-century Sweden**
Population N	80,808	459,591	271,130	8,201,968
Anchor N	14,034	79,895	56,947	1,419,282/ 3,428,225
Anchors/ Families (RS models)	9,447/ 2,186	68,724/ 12,205	56,663/ 14,746	1,408,177 / 884,975
Anchors/ Families (IS models)	9,447/ 2,186	61,493/ 11,940	56,010/ 14,708	363,744/ 200,000
Paternal age	35.23 (7.56)	36.28 (8.48)	34.37 (7.69)	31.84 (7.05)
Maternal age	31.53 (5.88)	29.58 (6.66)	31.54 (6.32)	28.34 (6.11)
Female/male infant mortality	11.1/12.9%	19.0/23.2 %	12.0/14.1 %	0.5/0.7%
Fertility	3.66 (2.89)	7.71 (4.57)	3.6 (3.17)	2.15 (1.11)

(married women)				
Male age at first child	29.29 (5.36)	27.92 (5.29)	28.13 (5.18)	28.07 (5.6)
Male age at last child	39.6 (7.5)	44.19 (8.59)	37.52 (8.29)	33.57 (6.14)

Table 1. Descriptive statistics. RS: reproductive success. IS: infant survival. Numbers in parentheses are standard deviations. Years refer to the birth years of the anchors. For 20th-century Sweden, fertility-related numbers are from 1947-1959 (first N given) and mortality numbers are from 1969-2000 (second N given).

The first population are inhabitants of the Krummhörn in contemporary Germany [32]. They were quite isolated and had a stable population size. We focussed on the 14,034 anchors born between 1720 and 1835. Married female anchors from this period had on average 3.7 children.

The second population are the French settlers of the Saint-Lawrence valley in contemporary Québec, Canada [33,34]. They were an isolated frontier population in a harsh climate but they also had access to abundant resources and unsettled land. We focussed on the 79,895 anchors born between 1670 and 1740. Married female anchors from this period had on average 7.7 children. In this dataset, we had access to deep pedigrees, allowing us to compare not only siblings for paternal age, but also cousins for grandpaternal age in a within-extended-family design.

The third population are Swedes in the Sundsvall, Northern inland (Karesuando to Undersåker, includes Sami people), Linköping and Skellefteå regions [35,36]. All individuals in Skellefteå and most individuals in Sundsvall

were linked between church parishes. In the other regions, some individuals appeared in more than one parish. We focussed on the 56,947 anchors born between 1737 and 1850. Married female anchors from this period had on average 3.6 children.

Our modern data is the whole population of Sweden. The Swedish Multi-Generation Register includes records of individuals born after 1932 and alive by 1962, as well as their parents. The dataset was linked to the Cause of Death register that includes death dates. Information about marriages was derived from the population register and the Longitudinal Integration Database for Health Insurance and Labour Market Studies [37]. Individuals who ever had the civil status of married, widowed or divorced were counted as *ever married*. Because of censoring in this dataset, we focussed on the 1,419,282 anchors born between 1947 and 1959 for reproductive outcomes and the 3,428,225 anchors born between 1969 and 2000 for survival outcomes. Ever married female anchors from the earlier period had on average 2.2 children (never married: 1.1). Hormonal contraception was widely available to and used by anchors born between 1947 and 1959.

Statistical approach

We employed generalized mixed effect regressions with a group-level effect per family to compare full biological siblings within families. We used the R package brms [38] to fit Bayesian regression models using the probabilistic programming language Stan [39], and adjusted for average paternal age within families to isolate the effect of paternal age differences

between siblings. We adjusted for birth cohort in five-year groupings (small groupings at the edge of the range were lumped) to account for secular changes in mortality and fertility, as well as residual censoring. We adjusted for parental deaths in the first 45 years of life to remove effects related to orphanhood and parental senescence (0-1, 2-5, 6-10, …, 45+, unknown). We adjusted for maternal age (up to 20, 21-34, 35+), which we *binned* to reduce multicollinearity with paternal age and to capture nonlinear effects. We also adjusted for number of siblings, number of older siblings (0-5, 5+), and being born last. We used weakly informative priors that are documented in detail in the online supplement. The modelling assumptions reflected herein were tested for robustness, as documented below.

We analysed reproductive success for *all* offspring, including those who died in childhood or never married. We used a two-process hurdle-Poisson family with a log link. In such a model, zeroes in the outcome variable are modelled as arising from a different process, e.g. not clearing the *hurdle* of survival and marriage before attempting reproduction. In the 20th-century Swedish data, we fitted a simpler Poisson model because child mortality was very low.

We separated effects into four successive episodes of natural and sexual selection. To separate the episodes, we adjusted for success in the preceding episode. *e1* survival of the first year, *e2* survival until age 15 conditional on *e1* survival of the first year, *e3* marriage conditional on *e2*, and *e4* number of children, conditional on *e3*. For *e4,* we included only ever-married anchors

and adjusted for their number of spouses. In 20th-century Sweden, we also examined *e5* divorce, conditional on *e3*, even though this is arguably not clearly an episode of selection. All models were fit using a Bernoulli regression with a cauchit link to decrease the influence of extreme values [40], except *e4* which was fit using a Poisson regression with a log link. In 20th-century Sweden, we could not fit our survival models to the whole available dataset for computational reasons and hence used a randomly drawn subset (~10% of the 3.4m available).

We used approximate leave-one-out cross-validation [41] as implemented in brms to compare four models: *m1* with a linear effect of paternal age, without the group-level effect for family, *m2* without a paternal age effect, but with the group-level effect, *m3* like *m2* but with a linear paternal age effect, and *m4,* like *m3,* but additionally with a thin-plate spline smooth [42] on the paternal age effect to capture nonlinearity. Comparing *m1* and *m3* allows us to assess the usefulness of group-level effects, comparing *m2* and *m3* we test whether the inclusion of paternal age improves the model fit, comparing *m3* and *m4,* we test the paternal age effect for nonlinearity.

After this, we ran several robustness checks to test the modelling assumptions in our main models, using *m3* as the baseline model. We carried out the following analyses: *r1* relaxed exclusion criteria (not in 20th-century Sweden), *r2* had only birth cohort as a covariate, *r3* adjusted for birth order continuously, *r4* adjusted for number of dependent siblings (younger than 5, alive at anchor birth) instead of birth order, *r5* interacted birth order with

number of siblings, *r6* did not adjust for birth order, *r7* adjusted only for parental loss in the first 5 years, *r8* adjusted for being the first- or last-born adult son, *r9* adjusted for a continuous nonlinear thin-plate spline smooth [42] for birth year instead of 5-year bins, *r10* added a group-level slope for paternal age, *r11* included separate group-level effects for each parent instead of one per marriage, *r12* added a moderation by anchor sex, *r13* adjusted for paternal age at first birth, *r14* compared a model with linear group fixed effects, *r15* added a moderator by region and group-level effects by church parish (not in 20th-century Sweden), *r16* was restricted to the region Skellefteå (only in historical Sweden), *r17* tested whether hypothetical cases of Down's syndrome could explain the effects, *r18* reversed hurdle Poisson and Poisson distribution for the respective populations, *r19* assumed a normal distribution for the outcome, *r20* did not adjust for maternal age, *r21* adjusted for maternal age continuously, *r22* relaxed exclusion criteria and included 30 more years of birth cohorts, allowing for more potential censoring, *r23* used different weakly informative priors, *r24* used noninformative priors (comparable with maximum likelihood), *r25* controlled for migration status (not in 20th-century Sweden), *r26* separated parental age contributions (only in 20th-century Sweden). More detailed descriptions of all robustness analyses can be found in the supplement section 6.2, code and detailed results are on the online supplementary website [31].

For the 20th-century Sweden data, we used a random subset of 80,000 families in the robustness analyses for computational reasons. We reran

analyses with all data if the paternal age effect deviated strongly from the *m3* estimate.

We also ran two sensitivity analyses to test whether results could be explained by late-life mortality or reproductive timing of the anchors. To contextualize contemporary reproductive timing trends, we also compared reproductive timing across populations.

Effect sizes were calculated as the median effect estimate of a 10-year increase in paternal age with a 95% credibility interval.

Results

In our main model *m3*, we found negative effects of paternal age on anchor's number of children in all four populations: a decrease per decade of paternal age of -3.0% (95% credibility interval: [-6.1,0.2] in Québec, -3.4% [-5.9,-0.9] in 20th-century Sweden, -7.3% [-13.4,-1.1] in historical Sweden, and -8.4% [-24.8,12.0] in the Krummhörn. These effects appeared to be fairly linear in *m4* (Figure 1), although visual inspection and approximate leave-one-out cross-validation [41] showed the effect tapering off after age 45 in 20th-century Sweden (~4% of children were born to fathers older than 45, see S.5.4.5.1) and after age 50 in Québec in (~8% of children, see S.3.4.5.1). In historical Sweden, paternal age had a slight positive effect in *m1* before using sibling comparisons, in the other populations the effect was negative in all models. In the Krummhörn population, the effects of birth order, maternal and paternal age could not be disentangled well, as credible intervals were very wide when

these covariates were considered together. Credible intervals (95%) for paternal age excluded zero for *m3* in both Swedish populations and for *m4* in Québec and 20th-century Sweden. These main models are detailed in the supplement sections 2-5.

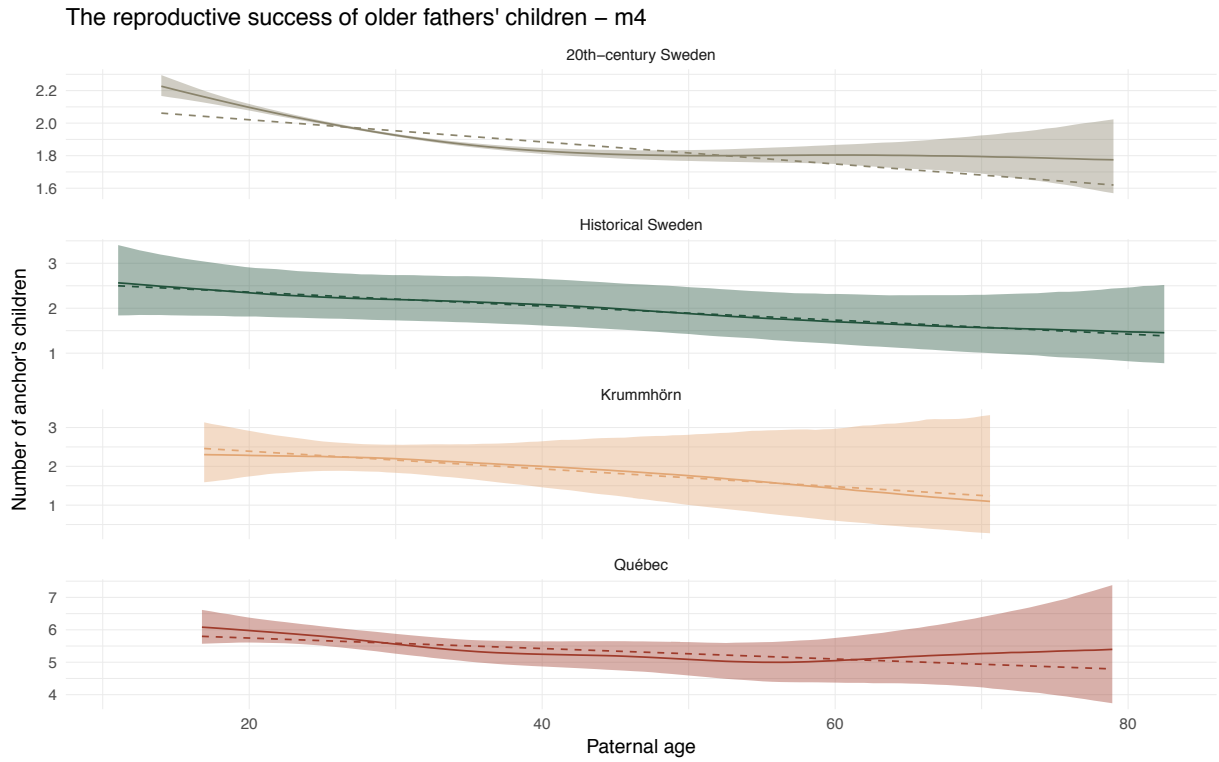


Fig. 1: Paternal age effects on number of surviving children.

Marginal effect plots for paternal age effect splines estimated in *m4*. Covariates were set to their mean or reference level, respectively. The solid lines show the posterior median; the dashed line is a linear line fit over the spline and inversely weighted by standard error to examine whether the spline fit deviates from linearity. The shaded areas show the 95% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.

In our selective episode analyses (Figure 2), we consistently found small negative associations between paternal age and anchor's survival to the first

year of life in the pre-industrial populations (*e1*). Comparing children of 25- and 35-year-old fathers, yielded percentage decreases of -2.1 (95% credible interval [-0.2,-5.4]), -1.0 [-0.7,-1.5], and -1.8 [-1.1,-3.1] in the Krummhörn, Québec and historical Sweden respectively. In the 20th-century Swedish population, infant mortality was very low, and the effect size of paternal age on infant survival, though negative, was correspondingly small (-0.05 [-0.03,-0.06]). Survival to age 15 years (*e2*) was not associated with paternal age (effects ranging from -0.2 to 0.1). Probability of ever marrying (*e3*) was inconsistently associated with paternal age, negatively in the Krummhörn population (-5.2), positively in historical Sweden (7.9), with negligible associations in Québec and modern Sweden (0.0 and 0.8), and the association in historical and 20th-century Sweden turned negative when not accounting for parental loss (not shown). Number of children (*e4*), after accounting for marriage success, was negatively associated with paternal age in 20th-century Sweden (-3.8 [-4.6;-3.0]) and historical Sweden (-5.4 [-8.9;-1.6]), but non-robustly positively associated in the Krummhörn population (15.62, negatively when not adjusting for birth order, not shown) and negligibly associated in Québec (0.9 [-1.3; 3.2]). Paternal age did not predict probability of divorce in 20th-century Sweden (-0.3 [-0.78;0.17]).

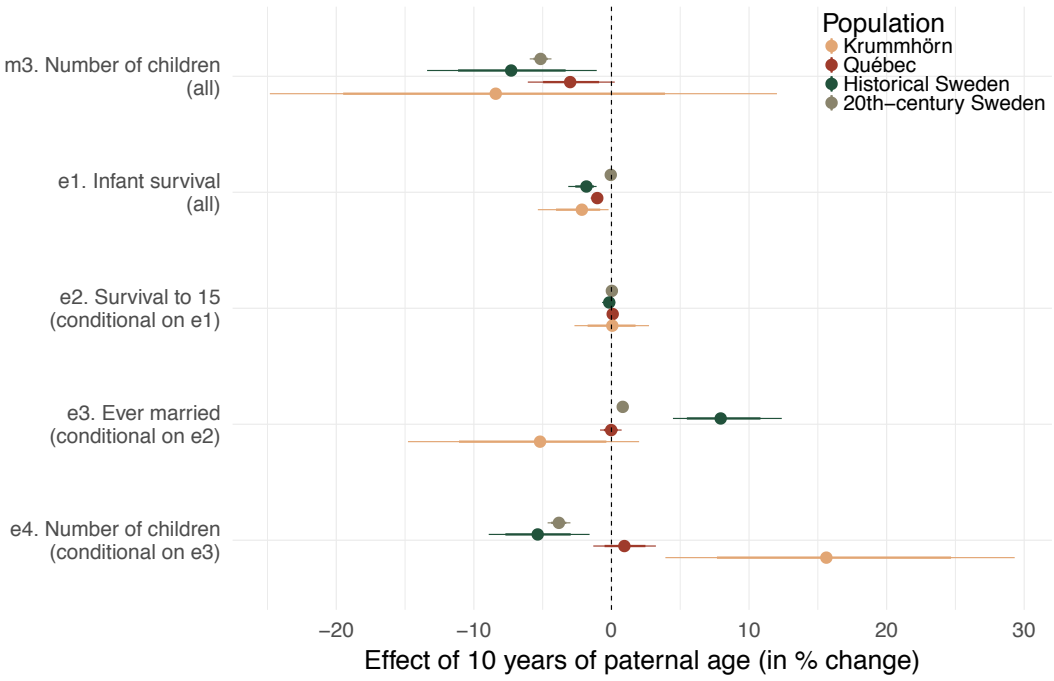


Fig. 2: Paternal age effects on subsequent selective episodes.

Estimated percentage changes in the respective selective episode (comparing children of 25- to 35-year-old fathers) with 80% and 95% credibility intervals.

In the grandpaternal age analyses in Québec, we found negative effects of both the paternal and maternal grandfather's age, that were roughly equal in size (paternal grandfather: -7% [-4,-9%], maternal grandfather: -5% [-2,-8%] fewer children).

In our robustness analyses (Figure 3), estimated paternal age effect sizes varied with our modelling assumptions. The paternal age effect was negative throughout almost all models in the two Swedish populations, and varied more widely in the Québec and Krummhörn models. In the Krummhörn, only the simplest model *r2* clearly supported a negative paternal age effect, but across robustness checks the estimate tended to be negative.

In our sensitivity analyses, we found mortality could mostly account for any paternal age effects on reproductive success in the two non-Swedish populations, but not in the Swedish populations. Among those who ever reproduced, paternal age did not predict reproductive success after accounting for anchor's age at first and last birth (confer supplement [31]).

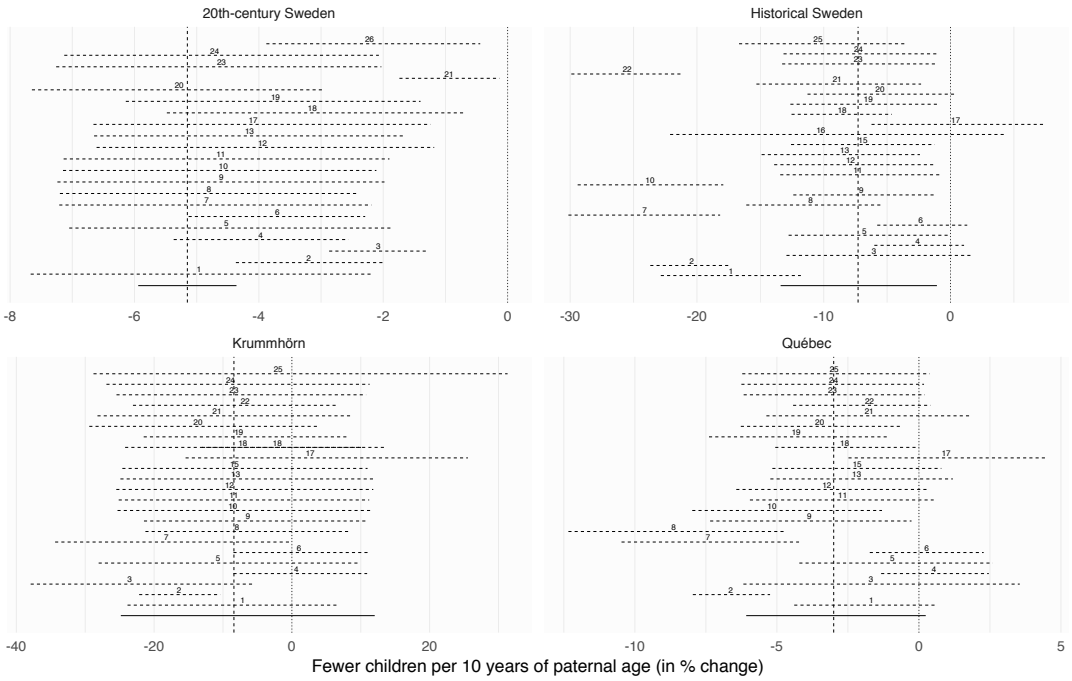


Fig. 3 Robustness checks across 26 models

Estimates of the effect of a ten-year difference in paternal age on number of children from model *m3* and up to 26 variations on this basic model (described in the method section and in further detail on the supplementary website). The horizontal dashed and solid lines show 95% credibility intervals. The point and vertical dashed line show the estimate from *m3*. The distance of the numbers to the vertical dashed line shows how much estimates can vary depending on the model specification. Estimates for the analyses in 20th-century Sweden are based on a subset of the data for computational reasons (except models m3, r3, r21, and r26).

Further details, including effect sizes and marginal effect plots for all covariates, model summaries, and R code for each of the models can be found on the online supplementary website at https://rubenarslan.github.io/paternal_age_fitness/ [31].

Discussion

We found robust evidence for negative paternal age effects on reproductive success in all four populations. Results held up after adjusting for numerous covariates, that capture alternative non-genetic explanations, including offspring sex, birth cohort, number of siblings, number of older siblings, maternal age, and loss of either parent up to age 45, and after checking robustness across 26 alternative models. In historical Sweden, a slight positive effect turned negative after we used sibling comparisons, showing that systematic confounding between reproductive timing and unobserved familial characteristics could obscure an effect. In all populations, effects were consistent with a roughly linear dose-response relationship between paternal age and number of children. Effects were largest in the Krummhörn (although estimates were uncertain in this smallest population), followed by historical Sweden, and similarly sized effects in Québec and 20th-century Sweden. These differences seemed to be mainly driven by differences in the first selective episode, survival of the first year. The 95% credibility intervals for all effect sizes overlapped across populations.

Even across three generations, we found negative grandpaternal age effects on offspring reproductive success for both grandfathers in Québec.

When we separately examined the selective episodes along the lifespan, paternal age effects on survival to the first year were negative across all historical populations (-1% in Québec to -2% in the Krummhörn and historical Sweden), but negligibly small in 20th-century Sweden (-0.05%). We found no robust pattern of effects on survival to age 15 and the odds of getting married. Some selective episode effects changed substantially depending on certain covariates, which may result from adjusting for a collider, mediator, or highly collinear variable. Therefore, we advocate only cautious interpretation of the analyses where the estimate changed substantially upon removal of a covariate, especially in the Krummhörn. In the Swedish populations, the number of children was negatively associated with paternal age after adjusting for marriage success and survival to age 15. Consistent with this, our sensitivity analyses showed that mortality could not explain the paternal age effect in the Swedish populations. This may, however, reflect a mere difference in statistical power to detect remaining effects, as opposed to a substantive difference between populations.

In 20th-century Sweden, the effect in the last selective episode, on number of children, was much stronger than the effect on infant mortality. Infant mortality in Sweden is among the lowest in the world. Because more than 99% of children brought to term in the years 1969 to 1999 survived, there is little room for selection during this selective episode. Future research should examine whether conditions that used to cause infant mortality, such as preterm birth, are simply no longer harmful thanks to advances in peri- and postnatal care, or whether selection has been partially displaced to before

birth or to later in life. We might expect displaced selection to take place before birth in some cases, as abortions end one fifth of all known pregnancies in Western Europe [43]. Most are elective, not therapeutic [44], but even women electing to have an abortion may do so selectively after considering their own age and paternal characteristics, including age [45]. Some paternal-age-linked conditions such as developmental disorders [4] might be detected in prenatal screening. Some diseases that would have led to early death in our historical populations might also put the afflicted at a disadvantage in later episodes of selection in 20th-century Sweden, e.g. people with paternal-age-associated [4] developmental disorders might be less likely to marry and have children.

We tried to adjust for all non-biological explanations that could be modelled using our data. Still, it is possible that e.g. parental investment declines with paternal age in such a manner that our adjustments for parental loss, mother's age, birth order and various other covariates in our robustness analyses could only insufficiently correct for this. Such residual confounding might lead to inflated estimates of any biological paternal age effect.

Moreover, several non-genetic biological explanations for paternal age effects have been suggested in the literature. Eisenberg et al. [46] linked advanced paternal age to longer offspring telomeres, but it remains unclear whether this association is causal, whether it would differ between siblings and whether it could mediate phenotypic effects. Some authors [47,48] have also speculated that advanced paternal age might lead to errors in epigenetic

regulation or might be linked to imprinting. Because preimplantation embryos undergo extensive demethylation and reprogramming [49,50], such transgenerational effects are controversial. Still, researchers [51–53] have searched for associations between paternal age and the methylation of certain genes in sperm and foetal cord blood. The use of small, clinical samples renders early work hard to generalise, but some associations have been reported.

Maternal age is another matter: its effects on aneuploidies are well established in the literature [54]. Although we adjusted for maternal age effects, parents' ages within families increase in lockstep. Their effects are thus difficult to separate in the largely pre-industrial monogamous populations. Even though maternal age is linked to aneuploidies, most aneuploid conceptions are not carried to term and even live-born children rarely get old. Only children with Down's syndrome live longer, but they are rarely fertile. Our robustness checks suggest Down's syndrome cannot fully explain the reported effects. In modern epidemiological data, specific syndromes could be easily excluded to test their contribution. Recent studies also estimated small effects of maternal age on single nucleotide *de novo* mutations [4,7]. Better understanding the mechanisms by which parental age is linked to offspring outcomes therefore seems to be a more worthwhile and achievable goal than perfectly separating each parent's contribution. Still, in modern Sweden we could separate parents' ages better, and in our robustness analyses paternal age still negatively predicted number of children after accounting for maternal

age continuously, the average parental age for each parent, and a dummy variable for teenage mothers.

Apart from these substantive alternative explanations, we also considered several methodological concerns. First and foremost, the highly collinear covariates maternal age, birth order and parental loss made it difficult to separate their contributions from that of paternal age. Standard errors were wide and different defensible operationalisations resulted in non-negligible effect size changes in our robustness analyses. Previous work rarely adjusted for parental loss to the extent that we did. This adjustment is debatable, because parental death can be both a cause and a consequence of offspring death. Still, from our robustness checks, we concluded that adjusting for parental loss is usually sensible and results of such adjustments should be reported in future work. Birth order, on the other hand, had little effect in most of our models, but adjusting for it often led to an increase of the paternal age effect size. Second, our church record data in particular have some shortcomings. Some children who died before baptism may have gone unrecorded, death records may be missing, and migration might lead to unobserved censoring [55]. Fortunately, judging from the consistency of our robustness analyses, it is at least plausible that these problems are unrelated to paternal age after adjusting for covariates in our models, and we assume that by using four different populations we limited bias.

After all these adjustments, we still found negative paternal age effects on several measures of evolutionary fitness across populations. But what can

explain these effects? The work of Kong et al. and others [1,6] has demonstrated a strong and likely causal effect of paternal age on de novo genetic mutations, but it is not clear that the paternal age effects reported here and in the literature are driven predominantly by de novo mutations [56]. One approach is to adjust for confounders, as we discuss above. Another is to derive expected effect size estimates from evolutionary genetic calculations. Gratten et al. [56] made the point that many reported paternal age effects in the psychiatric literature are implausibly large and calculated plausible effect sizes for mutational components of paternal age effects. Hayward et al. [22] estimated a paternal age effect on fitness components and attempted to compare their effect size to published estimates of the genome-wide deleterious mutation rate per generation (U) [3] times the mean selection effect against a deleterious mutation ($\bar{h}s$), yielding the estimated mutation-caused decrease in fitness as a percentage [27]. As paternal age does not perfectly predict the number of *de novo* mutations per generation, any estimate of paternal age effects on fitness would be expected to be slightly lower than $U\bar{h}s$. Unfortunately, no mean selection effect has been estimated for non-coding mutations yet and many unknowns and approximately-knowns enter the equation for estimates of the genome-wide deleterious mutation rate. Thus, only a range of plausible values can be drawn from the literature. Hayward et al. estimated values for $U\bar{h}s$ based on only nonsynonymous mutations ranging from 0.016-0.031 [22,27,57]. Estimates including mutations at all functional sites are even less certain; 0.11-0.22 are high estimates based on assuming the same mean selection as against deleterious

nonsynonymous mutations. If we now assume an increase of 2 mutations per year of paternal age [1] and estimate the per-generation decline in fitness from *de novo* mutations by comparing the child of an average father aged 30 years, transmitting 60 mutations, with the child of a hypothetical father transmitting no mutations, for our models *m3* in all four populations, we obtain 0.16, 0.07, 0.20, and 0.14 in the Krummhörn, Québec, historical and 20th-century Sweden respectively. Using the arguably better estimate from our robustness analysis *r26* in which we could better adjust for maternal age in 20th-century Sweden, we obtain an estimate of 0.065. Given the imperfect correlation between paternal age and *de novo* count, the variability of estimates in our robustness checks, sampling error and the plausibility of residual confounding, we think our estimates are on the high side of the real value, but not completely at odds with Hayward et al.'s calculations of $U\bar{h}s$ and consistent with their own estimated value of 0.12. We have also explored the relevant parameter space from Gratten et al. [56] and found the resulting effect sizes broadly consistent with the results from our infant survival models. These plausibility checks are documented in greater detail in the online supplement [31].

Implications and conclusions

Across four large population-based datasets, we found robust support for the prediction that higher paternal age linearly decreases offspring fitness. Although we cannot be sure that we succeeded in isolating an effect of *de novo* mutations given the multiple alternative explanations and methodological caveats, the effects are detectable in all four populations and hence plausibly

caused to some extent by paternal age. Depending on their cause, but not only if that cause is mutational, paternal age effects could have implications for policy: Descriptive data show a fall from 1930 to 1970 and a steady rise in maternal and paternal ages since 1970 in Sweden. However, average parental ages in 2010 were still lower than in 1737-1880 (supplement section 7). Although people start reproducing later, they also stop earlier. Contrary to common news and lay scientific accounts, contemporary parents do not reproduce unprecedentedly late *on average* [1,45,58]. While advanced parental ages at *first* birth may entail smaller families, pre-industrial populations had similar average ages at birth and were not overwhelmed by mutational stress. So, we do not predict that contemporary reproductive timing will lead to unprecedented or unbearable *de novo* mutational loads and concomitant changes in the prevalence of genetic disorders. The decline in fitness with paternal age suggests that purifying selection is still effective in a modern population with hormonal contraception, social transfers, and modern medicine. This runs counter to oft-repeated predictions of mutational doom by relaxed selection [3,59–61].

Although our design is not ideal for separating the influence of maternal and paternal age, many secular trends and policies will affect both. Future research could use genome-sequenced families with functionally annotated and phased mutations to better characterize the contribution of paternal age [4]. Future research could also isolate a biological paternal age effect on early mortality in nonhuman animals with large recorded pedigrees, such as artificially inseminated breeding cattle. This would rule out most social

confounds by design, but the much shorter breeding lifespan might limit generalizability to humans.

Ethics Statement

For integrity reasons, national registration numbers are replaced with unique sequence numbers when register data are used for research purposes. Ethical approval was given by the regional ethics committee in Stockholm (dnr. 2009/939-31/5). Data from the pre-industrial populations was drawn from church records and does not include living individuals.

Competing interests

We have no competing interests.

Authors' contributions

RCA and LP conceived of the study. RCA coordinated it, carried out the analyses and drafted the manuscript. KPW provided guidance and pre-processing for church record data, and replicated central analyses in Stata. PCB wrote *brms* and provided guidance for data analysis and interpretation. EMF and CA contributed the contemporary Swedish data. EMF also provided guidance and pre-processing. EV contributed the Krummhörn data. KPW, KJHV, BPZ, MM, CA, and LP helped design the study, interpret the data and critically revised the manuscript. All authors helped draft the manuscript and gave final approval for publication.

Data availability

Because of identification concerns and licensing restrictions, the data are only available from the dataset maintainers and underlie some usage restrictions.

Acknowledgements

RCA thanks Jarrod Hadfield, Ben Bolker and Holger Sennhenn-Reulen for their statistical advice, Silvia Bradatsch for proofreading the online supplement and Claude Bhérer, Peter Visscher and five anonymous reviewers for their helpful comments on earlier versions of this manuscript. MM acknowledges the support of the European Research Council grant 336475. CA and EMF acknowledge financial support from the Swedish Research Council through the Swedish Initiative for Research on Microdata in the Social And Medical Sciences (SIMSAM) framework grant no. 340-2013-5867.

References

1. Kong A *et al.* 2012 Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475. (doi:10.1038/nature11396)

2. Shendure J, Akey JM. 2015 The origins, determinants, and consequences of human mutations. *Science* **349**, 1478–1483. (doi:10.1126/science.aaa9119)

3. Keightley PD. 2012 Rates and Fitness Consequences of New Mutations in Humans. *Genetics* **190**, 295–304. (doi:10.1534/genetics.111.134668)

4. Deciphering Developmental Disorders Study. 2017 Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438. (doi:10.1038/nature21062)

5. Tomasetti C, Vogelstein B. 2015 Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81. (doi:10.1126/science.1260825)

6. Ségurel L, Wyman MJ, Przeworski M. 2014 Determinants of mutation rate variation in the human germline. *Annu. Rev. Genomics Hum. Genet.* **15**, 47–70. (doi:10.1146/annurev-genom-031714-125740)

7. Wong WSW *et al.* 2016 New observations on maternal age effect on germline de novo mutations. *Nat. Commun.* **7**, 10486. (doi:10.1038/ncomms10486)

8. Kloosterman WP *et al.* 2015 Characteristics of de novo structural changes in the human genome. *Genome Res.* (doi:10.1101/gr.185041.114)

9. Reichenberg A *et al.* 2006 Advancing paternal age and autism. *Arch. Gen. Psychiatry* **63**, 1026–1032. (doi:10.1001/archpsyc.63.9.1026)

10. D'Onofrio BM, Rickert ME, Frans E, Kuja-Halkola R, Almqvist C, Sjölander A, Larsson H, Lichtenstein P. 2014 Paternal age at childbearing and offspring psychiatric and academic morbidity. *JAMA Psychiatry* (doi:10.1001/jamapsychiatry.2013.4525)

11. Iossifov I *et al.* 2014 The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221. (doi:10.1038/nature13908)

12. Zietsch BP, Kuja-Halkola R, Walum H, Verweij KJH. 2014 Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population. *Proc. Natl. Acad. Sci.* , 201310058. (doi:10.1073/pnas.1310058111)

13. Arslan RC, Penke L, Johnson W, Iacono WG, McGue M. 2014 The effect of paternal age on offspring intelligence and personality when controlling for parental trait levels. *PLoS ONE* **9**, e90097. (doi:10.1371/journal.pone.0090097)

14. Frans E, MacCabe JH, Reichenberg A. 2015 Advancing paternal age and psychiatric disorders. *World Psychiatry* **14**, 91–93. (doi:10.1002/wps.20190)

15. Myrskylä M, Silventoinen K, Tynelius P, Rasmussen F. 2013 Is later better or worse? Association of advanced parental age with offspring cognitive ability among half a million young Swedish men. *Am. J. Epidemiol.* **177**, 649–655. (doi:10.1093/aje/kws237)

16. Byars SG, Stearns SC, Boomsma JJ. 2014 Opposite risk patterns for autism and schizophrenia are associated with normal variation in birth size: phenotypic support for hypothesized diametric gene-dosage effects. *Proc R Soc B* **281**, 20140604. (doi:10.1098/rspb.2014.0604)

17. Houle D, Hughes KA, Hoffmaster DK, Ihara J, Assimacopoulos S, Canada D, Charlesworth B. 1994 The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. *Genetics* **138**, 773–785.

18. Nussey DH, Froy H, Lemaitre J-F, Gaillard J-M, Austad SN. 2013 Senescence in natural populations of animals: Widespread evidence and its implications for bio-gerontology. *Ageing Res. Rev.* **12**, 214–225. (doi:10.1016/j.arr.2012.07.004)

19. Schroeder J, Nakagawa S, Rees M, Mannarelli M-E, Burke T. 2015 Reduced fitness in progeny from old parents in a natural population. *Proc. Natl. Acad. Sci.* (doi:10.1073/pnas.1422715112)

20. Fieder M, Huber S. 2014 Paternal age predicts offspring chances of marriage and reproduction. *Am. J. Hum. Biol.* , 339–343. (doi:10.1002/ajhb.22644)

21. Gavrilov LA, Gavrilova NS. 1997 Parental age at conception and offspring longevity. *Rev. Clin. Gerontol.* **7**, 5–12. (doi:10.1017/S0959259897000026)

22. Hayward AD, Lummaa V, Bazykin GA. 2015 Fitness consequences of advanced ancestral age over three generations in humans. *PLoS ONE* **10**, e0128197. (doi:10.1371/journal.pone.0128197)

23. Zhu JL, Vestergaard M, Madsen KM, Olsen J. 2008 Paternal age and mortality in children. *Eur. J. Epidemiol.* **23**, 443–447. (doi:10.1007/s10654-008-9253-3)

24. Hayward AD, Nenko I, Lummaa V. 2015 Early-life reproduction is associated with increased mortality risk but enhanced lifetime fitness in pre-industrial humans. *Proc. R. Soc. Lond. B Biol. Sci.* **282**, 20143053. (doi:10.1098/rspb.2014.3053)

25. Gillespie DOS, Russell AF, Lummaa V. 2013 The effect of maternal age and reproductive history on offspring survival and lifetime reproduction in preindustrial humans. *Evolution* **67**, 1964–1974. (doi:10.1111/evo.12078)

26. Myrskylä M, Elo IT, Kohler IV, Martikainen P. 2014 The association between advanced maternal and paternal ages and increased adult mortality is explained by early parental loss. *Soc. Sci. Med.* **119**, 215–223. (doi:10.1016/j.socscimed.2014.06.008)

27. Eyre-Walker A, Keightley PD. 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618. (doi:10.1038/nrg2146)

28. Frans E. 2013 *High paternal age and risk of psychiatric disorders in offspring*. Stockholm, Sweden: Karolinska Institutet. See http://hdl.handle.net/10616/41713.

29. Gagnon A, Heyer E. 2001 Intergenerational correlation of effective family size in early Quebec (Canada). *Am. J. Hum. Biol.* **13**, 645–659.

30. Austerlitz F, Heyer E. 1998 Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proc. Natl. Acad. Sci.* **95**, 15140–15144.

31. Arslan RC *et al.* 2017 Paternal age and offspring fitness: Online supplementary website https://rubenarslan.github.io/paternal_age_fitness/. *Zenodo* (doi:10.5281/zenodo.838961)

32. Voland E. 1990 Differential reproductive success within the Krummhörn population (Germany, 18th and 19th centuries). *Behav. Ecol. Sociobiol.* **26**, 65–72. (doi:10.1007/BF00174026)

33. Charbonneau H, Desjardins B, Légaré J, Denis H. 2000 The population of the St-Lawrence Valley, 1608–1760. *Popul. Hist. N. Am.* , 99–142.

34. Desjardins B. 1999 Le registre de la population du Québec ancien. *Ann. Démographie Hist.* **1998**, 215–226. (doi:10.3406/adh.1999.1946)

35. Edvinsson S. 2000 The Demographic Data Base at Umeå University—a resource for historical studies. *Handb. Int. Hist. Microdata Popul. Res.* , 231–248.

36. Demographic Data Base, Umeå University. 2014 Database description file. ID: U15006.

37. Ludvigsson JF, Almqvist C, Bonamy A-KE, Ljung R, Michaëlsson K, Neovius M, Stephansson O, Ye W. 2016 Registers of the Swedish total population and their use in medical research. *Eur. J. Epidemiol.* **31**, 125–136. (doi:10.1007/s10654-016-0117-y)

38. Bürkner P-C. in press brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.*

39. Carpenter B *et al.* 2015 Stan: a probabilistic programming language. *J. Stat. Softw.*

40. Koenker R, Yoon J. 2009 Parametric links for binary choice models: A Fisherian–Bayesian colloquy. *J. Econom.* **152**, 120–130.

41. Vehtari A, Gelman A, Gabry J. 2016 Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* (doi:10.1007/s11222-016-9696-4)

42. Wood SN. 2003 Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**, 95–114.

43. Sedgh G, Singh S, Shah IH, Åhman E, Henshaw SK, Bankole A. 2012 Induced abortion: incidence and trends worldwide from 1995 to 2008. *The Lancet* **379**, 625–632. (doi:10.1016/S0140-6736(11)61786-8)

44. Lohr PA, Fjerstad M, DeSilva U, Lyus R. 2014 Abortion. *BMJ* **348**. (doi:10.1136/bmj.f7553)

45. Blomberg Jensen M, Priskorn L, Jensen TK, Juul A, Skakkebaek NE. 2015 Temporal Trends in Fertility Rates: A Nationwide Registry Based Study from 1901 to 2014. *PLoS ONE* **10**, e0143722. (doi:10.1371/journal.pone.0143722)

46. Eisenberg DTA, Hayes MG, Kuzawa CW. 2012 Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10251–10256. (doi:10.1073/pnas.1202092109)

47. Perrin MC, Brown AS, Malaspina D. 2007 Aberrant epigenetic regulation could explain the relationship of paternal age to schizophrenia. *Schizophr. Bull.* **33**, 1270–1273. (doi:10.1093/schbul/sbm093)

48. Curley JP, Mashoodh R, Champagne FA. 2011 Epigenetics and the origins of paternal effects. *Horm. Behav.* **59**, 306–314. (doi:10.1016/j.yhbeh.2010.06.018)

49. Reik W, Dean W, Walter J. 2001 Epigenetic Reprogramming in Mammalian Development. *Science* **293**, 1089–1093. (doi:10.1126/science.1063443)

50. Schaefer CB, Ooi SKT, Bestor TH, Bourc'his D. 2007 Epigenetic Decisions in Mammalian Germ Cells. *Science* **316**, 398–399. (doi:10.1126/science.1137544)

51. Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT. 2014 Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. *PLoS Genet* **10**, e1004458. (doi:10.1371/journal.pgen.1004458)

52. Atsem S *et al.* 2016 Paternal age effects on sperm FOXK1 and KCNA7 methylation and transmission into the next generation. *Hum. Mol. Genet.* , ddw328. (doi:10.1093/hmg/ddw328)

53. Sharma R, Agarwal A, Rohra VK, Assidi M, Abu-Elmagd M, Turki RF. 2015 Effects of increased paternal age on sperm quality, reproductive outcome and associated epigenetic risks to offspring. *Reprod. Biol. Endocrinol.* **13**, 35. (doi:10.1186/s12958-015-0028-x)

54. Morris JK, Mutton DE, Alberman E. 2002 Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *J. Med. Screen.* **9**, 2–6. (doi:10.1136/jms.9.1.2)

55. Nault F, Desjardins B, Légaré J. 1990 Effects of reproductive behaviour on infant mortality of French-Canadians during the seventeenth and eighteenth centuries. *Popul. Stud.* **44**, 273–285.

56. Gratten J, Wray NR, Peyrot WJ, McGrath JJ, Visscher PM, Goddard ME. 2016 Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. *Nat. Genet.* **48**, 718–724. (doi:10.1038/ng.3577)

57. Eyre-Walker A, Woolfit M, Phelps T. 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900. (doi:10.1534/genetics.106.057570)

58. Willführ KP, Klüsener S. in prep. The evolution of mean paternal age from 1900 to the present day – are today's fathers really older than back in the days?

59. Kondrashov AS, Crow JF. 1993 A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2**, 229–234. (doi:10.1002/humu.1380020312)

60. Crabtree GR. 2012 Our fragile intellect. Part I. *Trends Genet.* , 1–3. (doi:10.1016/j.tig.2012.10.002)

61. Lynch M. 2016 Mutation and Human Exceptionalism: Our Future Genetic Load. *Genetics* **202**, 869–875. (doi:10.1534/genetics.115.180471)

Paternal age and offspring fitness

Contents

Title: Older fathers' children have lower evolutionary fitness across four centuries and in four populations Authors: Arslan, R. C., Willführ, K. P., Frans, E., Verweij, K. J. H., Bürkner, P.-C., Myrskylä, M., Voland, E., Almqvist, C., Zietsch, B. P., Penke, L. Proceedings of the Royal Society B. Biological Sciences.

See also full supplementary website: https://rubenarslan.github.io/paternal_age_fitness/.

1 Model description

All of the models described below have the following in common. Only the robustness check models deviate from this in the way described in section 6.

1.1 Estimation

We fit all models using `brms` v. 1.2.0, a Bayesian regression analysis statistical package. `brms` uses `Stan`, a probabilistic programming langugage to fit models using Hamiltonian Monte Carlo.

1.2 Covariates

We adjusted for average paternal age within families to isolate the effect of paternal age differences between siblings. We further adjusted for birth cohort in five-year groupings (small groupings at the edge of the range were lumped) to account for secular changes in mortality and fertility, as well as residual censoring. We adjusted for parental deaths in the first 45 years of life to remove effects related to orphanhood and parental senescence (in categories of 0-1, 2-5, 6-10, ..., 45+, unknown) for both parents separately. Parental loss at 45+ served as the reference category. We adjusted for maternal age (up to 20, 21-34, 35+), which we binned to reduce multicollinearity with paternal age and to capture nonlinear effects. A maternal age of 21-34 served as the reference category. We also adjusted for number of siblings continuously, number of older siblings (0-5, 5+), and being born last. Being first-born served as the reference category.

1.3 Model stratification

Except in model *m1*, we added group-level effects for each family (father-mother dyad) and then controlled for the average paternal age in the family. Hence, the effect of paternal age within families can be isolated from the effect between families. We are interested in the effect of paternal age within families, as this effect cannot be explained by e.g. genetic propensities of the father to reproduce later.

1.4 Priors

We used weakly informative normal priors with a standard deviation of 5 on the regression coefficients, Student's t priors with 3 degrees of freedom and a scale of 5 for the group-level standard deviations, and Student's t priors with 3 degrees of freedom and a scale of 10 for the splines.

1.5 Robustness tests

The modelling assumptions, including covariate choices and prior choices, reflected in the modelling approach above were tested for robustness, as documented in section 6 below.

2 Krummhörn

2.1 *m1*: No sibling comparison

Here, we ignore the pedigree structure of the data to see whether it matters for the estimation of the paternal age effect.

2.1.1 Model summary

Data: 9447 individuals.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born.

- **family**: hurdle_poisson
- **link**: log

2.1.2 Priors

prior	class
normal(0,5)	b

2.1.3 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1760-1765	1.00 [0.90;1.12]	0.97 [0.72;1.33]
birth cohort 1765-1770	0.88 [0.80;0.97]	0.76 [0.58;0.99]
birth cohort 1770-1775	0.90 [0.82;0.99]	0.94 [0.73;1.23]
birth cohort 1775-1780	0.98 [0.89;1.07]	0.83 [0.64;1.07]
birth cohort 1780-1785	0.88 [0.80;0.97]	0.78 [0.60;1.00]
birth cohort 1785-1790	0.91 [0.83;1.00]	0.68 [0.53;0.88]
birth cohort 1790-1795	0.94 [0.86;1.02]	0.75 [0.60;0.96]
birth cohort 1795-1800	0.90 [0.83;0.97]	0.66 [0.52;0.83]
birth cohort 1800-1805	0.89 [0.82;0.96]	0.60 [0.49;0.75]
birth cohort 1805-1810	0.87 [0.80;0.95]	0.77 [0.62;0.97]
birth cohort 1810-1815	0.91 [0.84;0.98]	0.66 [0.54;0.82]
birth cohort 1815-1820	0.86 [0.80;0.93]	0.52 [0.42;0.64]
birth cohort 1820-1825	0.83 [0.77;0.90]	0.61 [0.50;0.76]
birth cohort 1825-1830	0.82 [0.76;0.89]	0.61 [0.50;0.75]
birth cohort 1830-1835	0.84 [0.78;0.91]	0.60 [0.49;0.74]
Intercept	5.34 [4.76;6.03]	0.69 [0.50;0.96]
last born	0.95 [0.91;0.99]	1.08 [0.96;1.22]
male	1.08 [1.05;1.12]	1.30 [1.19;1.41]
maternal loss 0-1	1.11 [0.97;1.26]	4.51 [3.18;6.55]
maternal loss 1-5	1.00 [0.92;1.08]	1.77 [1.44;2.19]
maternal loss 10-15	1.04 [0.97;1.11]	1.59 [1.32;1.94]
maternal loss 15-20	1.00 [0.94;1.08]	1.34 [1.10;1.63]
maternal loss 20-25	1.01 [0.95;1.08]	1.30 [1.09;1.56]
maternal loss 25-30	0.99 [0.94;1.05]	1.20 [1.02;1.41]
maternal loss 30-35	0.96 [0.91;1.02]	1.22 [1.06;1.43]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
maternal loss 35-40	0.98 [0.92;1.03]	1.07 [0.93;1.23]
maternal loss 40-45	0.97 [0.92;1.03]	1.29 [1.10;1.52]
maternal loss 5-10	1.07 [1.00;1.14]	1.59 [1.31;1.93]
maternalage factor 14-20	0.93 [0.78;1.09]	1.30 [0.83;2.04]
maternalage factor 35-50	1.01 [0.96;1.06]	1.16 [1.02;1.31]
nr siblings	1.00 [0.99;1.01]	1.09 [1.06;1.11]
older siblings 1	1.03 [0.99;1.08]	1.03 [0.91;1.16]
older siblings 2	0.98 [0.93;1.03]	0.93 [0.80;1.07]
older siblings 3	0.97 [0.92;1.03]	0.93 [0.79;1.10]
older siblings 4	0.96 [0.90;1.03]	0.94 [0.78;1.13]
older siblings 5+	0.99 [0.92;1.06]	0.72 [0.59;0.89]
paternal loss 0-1	0.86 [0.76;0.98]	1.76 [1.28;2.43]
paternal loss 1-5	0.97 [0.89;1.06]	1.67 [1.33;2.11]
paternal loss 10-15	1.00 [0.94;1.07]	1.16 [0.96;1.38]
paternal loss 15-20	0.90 [0.85;0.96]	1.09 [0.92;1.31]
paternal loss 20-25	0.88 [0.82;0.94]	1.15 [0.96;1.36]
paternal loss 25-30	0.99 [0.93;1.05]	1.05 [0.89;1.25]
paternal loss 30-35	0.96 [0.90;1.01]	0.98 [0.84;1.16]
paternal loss 35-40	0.97 [0.92;1.03]	0.98 [0.84;1.15]
paternal loss 40-45	0.99 [0.92;1.05]	1.15 [0.95;1.38]
paternal loss 5-10	0.93 [0.87;1.00]	1.21 [1.00;1.49]
paternalage	1.00 [0.97;1.03]	1.09 [1.01;1.17]

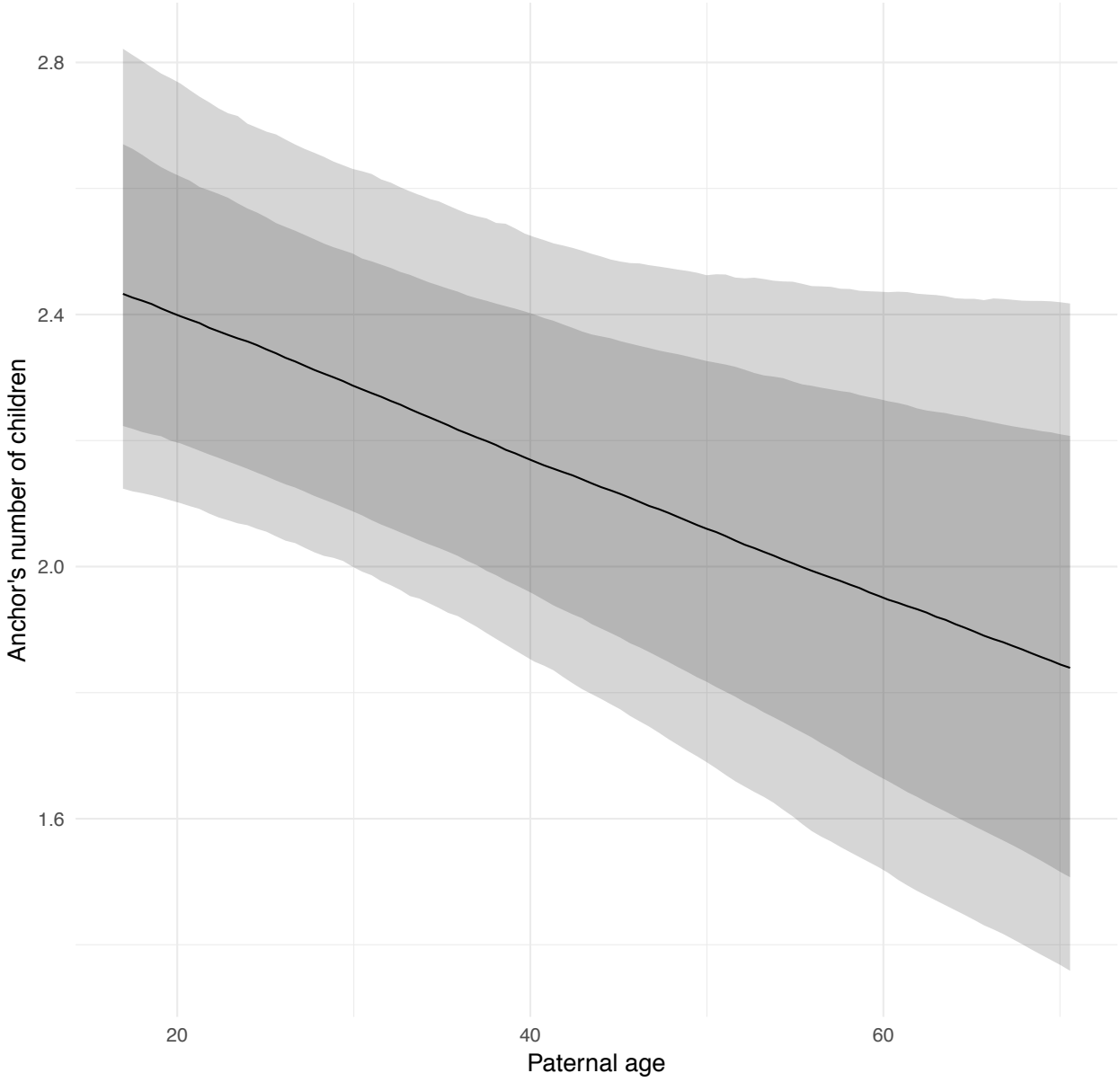
2.1.4 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-4.87	[-9.82; 0.17]	[-8.12;-1.57]

2.1.4.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



2.2 *m2*: Sibling comparison, no paternal age effect

Here, we compared siblings by including a random intercept for the family, but we modelled no effect for paternal age differences among siblings.

2.2.1 Model summary

Data: 9447 individuals nested in 2186 mother-father dyads.

Formula (Wilkinson notation): children ~ birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

2.2.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

2.2.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.48 [0.39;0.56]	0.23 [0.20;0.25]

2.2.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1760-1765	1.00 [0.89;1.13]	0.95 [0.68;1.32]
birth cohort 1765-1770	0.89 [0.80;1.00]	0.73 [0.55;0.97]
birth cohort 1770-1775	0.90 [0.80;1.00]	0.94 [0.71;1.25]
birth cohort 1775-1780	0.98 [0.89;1.09]	0.82 [0.61;1.08]
birth cohort 1780-1785	0.90 [0.81;1.01]	0.76 [0.57;1.02]
birth cohort 1785-1790	0.91 [0.82;1.01]	0.65 [0.49;0.86]
birth cohort 1790-1795	0.93 [0.84;1.03]	0.73 [0.56;0.96]
birth cohort 1795-1800	0.90 [0.82;1.00]	0.63 [0.49;0.80]
birth cohort 1800-1805	0.90 [0.82;0.99]	0.58 [0.45;0.74]
birth cohort 1805-1810	0.88 [0.80;0.97]	0.76 [0.60;0.97]
birth cohort 1810-1815	0.91 [0.83;1.00]	0.65 [0.51;0.82]
birth cohort 1815-1820	0.87 [0.80;0.95]	0.50 [0.39;0.62]
birth cohort 1820-1825	0.84 [0.77;0.91]	0.59 [0.47;0.74]
birth cohort 1825-1830	0.82 [0.75;0.89]	0.58 [0.46;0.72]
birth cohort 1830-1835	0.84 [0.76;0.92]	0.58 [0.46;0.73]
Intercept	5.15 [4.41;6.00]	0.69 [0.47;1.03]
last born	0.96 [0.91;1.00]	1.08 [0.96;1.23]
male	1.08 [1.04;1.12]	1.31 [1.20;1.43]
maternal loss 0-1	1.12 [0.96;1.30]	4.98 [3.49;7.17]
maternal loss 1-5	0.99 [0.90;1.08]	1.84 [1.46;2.33]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
maternal loss 10-15	1.03 [0.95;1.12]	1.62 [1.30;2.01]
maternal loss 15-20	1.01 [0.93;1.09]	1.38 [1.12;1.70]
maternal loss 20-25	1.01 [0.94;1.09]	1.31 [1.08;1.60]
maternal loss 25-30	0.98 [0.92;1.05]	1.22 [1.03;1.47]
maternal loss 30-35	0.95 [0.89;1.01]	1.25 [1.05;1.48]
maternal loss 35-40	0.97 [0.92;1.02]	1.08 [0.92;1.26]
maternal loss 40-45	0.97 [0.91;1.04]	1.32 [1.11;1.57]
maternal loss 5-10	1.08 [1.00;1.17]	1.64 [1.34;2.01]
maternalage factor 14-20	0.94 [0.79;1.13]	1.25 [0.81;1.98]
maternalage factor 35-50	1.01 [0.96;1.06]	1.20 [1.05;1.36]
nr siblings	1.00 [0.99;1.01]	1.08 [1.05;1.11]
older siblings 1	1.04 [0.99;1.09]	1.05 [0.92;1.20]
older siblings 2	0.98 [0.93;1.04]	0.96 [0.84;1.11]
older siblings 3	0.97 [0.91;1.03]	0.98 [0.83;1.15]
older siblings 4	0.97 [0.90;1.04]	1.00 [0.83;1.21]
older siblings 5+	0.99 [0.91;1.07]	0.78 [0.63;0.96]
paternal loss 0-1	0.86 [0.75;1.00]	1.84 [1.32;2.60]
paternal loss 1-5	0.97 [0.88;1.07]	1.75 [1.36;2.25]
paternal loss 10-15	1.01 [0.94;1.09]	1.19 [0.96;1.47]
paternal loss 15-20	0.91 [0.85;0.98]	1.12 [0.92;1.38]
paternal loss 20-25	0.89 [0.83;0.96]	1.18 [0.97;1.43]
paternal loss 25-30	0.99 [0.93;1.06]	1.07 [0.89;1.28]
paternal loss 30-35	0.97 [0.91;1.04]	0.99 [0.82;1.19]
paternal loss 35-40	0.99 [0.93;1.06]	0.99 [0.83;1.18]
paternal loss 40-45	0.99 [0.93;1.06]	1.16 [0.96;1.41]
paternal loss 5-10	0.94 [0.86;1.02]	1.24 [1.00;1.55]
paternalage mean	1.00 [0.96;1.03]	1.08 [0.99;1.18]

2.2.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

This model did not contain a within family paternal age predictor.

2.3 *m3*: Sibling comparison, linear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a linear effect for paternal age differences among siblings.

2.3.1 Model summary

Data: 9447 individuals nested in 2186 mother-father dyads.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

2.3.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

2.3.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.47 [0.39;0.56]	0.22 [0.20;0.25]

2.3.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1760-1765	1.00 [0.88;1.13]	0.95 [0.68;1.34]
birth cohort 1765-1770	0.89 [0.79;0.99]	0.73 [0.55;0.99]
birth cohort 1770-1775	0.89 [0.80;1.00]	0.94 [0.69;1.26]
birth cohort 1775-1780	0.98 [0.87;1.09]	0.82 [0.61;1.09]
birth cohort 1780-1785	0.90 [0.80;1.01]	0.76 [0.57;1.01]
birth cohort 1785-1790	0.91 [0.82;1.02]	0.66 [0.49;0.86]
birth cohort 1790-1795	0.93 [0.84;1.03]	0.73 [0.57;0.95]
birth cohort 1795-1800	0.90 [0.81;1.00]	0.63 [0.50;0.81]
birth cohort 1800-1805	0.89 [0.81;0.99]	0.58 [0.46;0.75]
birth cohort 1805-1810	0.88 [0.79;0.97]	0.76 [0.59;0.97]
birth cohort 1810-1815	0.91 [0.83;1.00]	0.65 [0.51;0.82]
birth cohort 1815-1820	0.87 [0.80;0.96]	0.49 [0.39;0.63]
birth cohort 1820-1825	0.83 [0.76;0.91]	0.59 [0.47;0.74]
birth cohort 1825-1830	0.81 [0.74;0.89]	0.58 [0.46;0.73]
birth cohort 1830-1835	0.83 [0.76;0.92]	0.57 [0.45;0.73]
Intercept	5.23 [4.49;6.07]	0.72 [0.49;1.05]
last born	0.96 [0.91;1.00]	1.08 [0.96;1.23]
male	1.08 [1.05;1.12]	1.32 [1.20;1.44]
maternal loss 0-1	1.10 [0.94;1.28]	4.85 [3.46;6.97]
maternal loss 1-5	0.98 [0.89;1.07]	1.80 [1.44;2.27]
maternal loss 10-15	1.03 [0.95;1.11]	1.60 [1.30;1.99]
maternal loss 15-20	1.00 [0.93;1.09]	1.37 [1.11;1.68]
maternal loss 20-25	1.00 [0.93;1.08]	1.30 [1.05;1.59]
maternal loss 25-30	0.98 [0.92;1.05]	1.21 [1.01;1.45]
maternal loss 30-35	0.95 [0.89;1.01]	1.24 [1.05;1.47]
maternal loss 35-40	0.97 [0.91;1.02]	1.07 [0.92;1.26]
maternal loss 40-45	0.97 [0.91;1.03]	1.32 [1.12;1.56]
maternal loss 5-10	1.07 [0.99;1.16]	1.62 [1.32;1.99]
maternalage factor 14-20	0.95 [0.79;1.13]	1.27 [0.80;2.00]
maternalage factor 35-50	1.00 [0.95;1.05]	1.14 [0.99;1.31]
nr siblings	1.01 [1.00;1.02]	1.11 [1.07;1.15]
older siblings 1	1.03 [0.97;1.08]	0.99 [0.85;1.15]
older siblings 2	0.95 [0.89;1.02]	0.86 [0.71;1.04]
older siblings 3	0.93 [0.84;1.02]	0.83 [0.65;1.06]
older siblings 4	0.91 [0.81;1.02]	0.80 [0.58;1.08]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
older siblings 5+	0.91 [0.78;1.05]	0.57 [0.38;0.85]
paternal loss 0-1	0.86 [0.74;0.99]	1.78 [1.26;2.55]
paternal loss 1-5	0.96 [0.87;1.06]	1.70 [1.32;2.21]
paternal loss 10-15	1.01 [0.93;1.09]	1.17 [0.95;1.45]
paternal loss 15-20	0.91 [0.84;0.98]	1.11 [0.92;1.35]
paternal loss 20-25	0.89 [0.82;0.96]	1.17 [0.96;1.41]
paternal loss 25-30	0.99 [0.92;1.06]	1.06 [0.88;1.27]
paternal loss 30-35	0.97 [0.91;1.04]	0.98 [0.82;1.17]
paternal loss 35-40	0.99 [0.93;1.05]	0.98 [0.83;1.18]
paternal loss 40-45	0.99 [0.92;1.06]	1.15 [0.95;1.39]
paternal loss 5-10	0.93 [0.86;1.02]	1.21 [0.97;1.51]
paternalage	1.07 [0.97;1.19]	1.30 [0.99;1.69]
paternalage mean	0.93 [0.83;1.04]	0.84 [0.64;1.11]

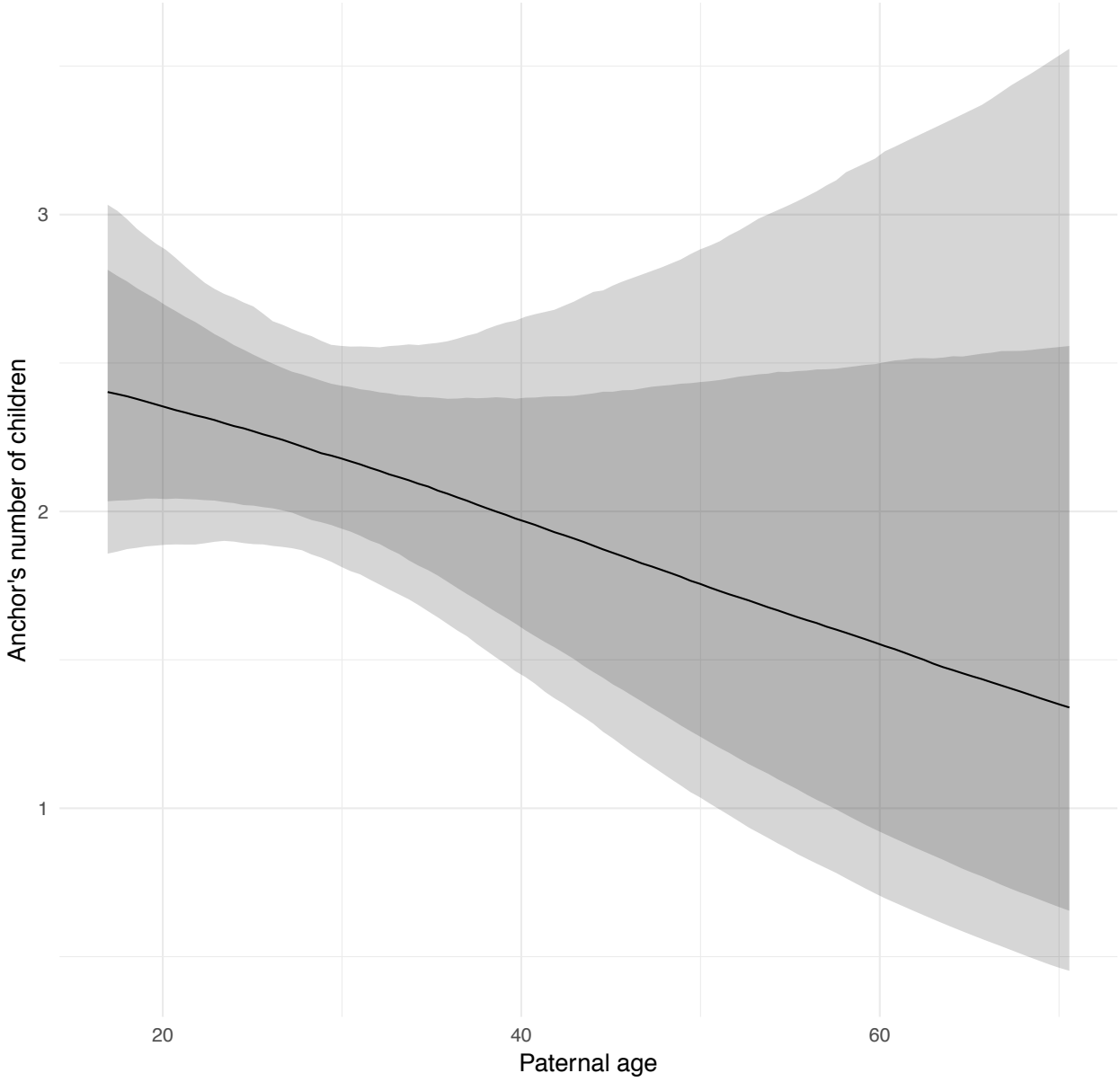
2.3.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-8.41	[-24.83; 12.03]	[-19.50; 3.89]

2.3.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



2.4 *m4*: Sibling comparison, nonlinear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a possibly nonlinear effect for paternal age differences among siblings.

2.4.1 Model summary

Data: 9447 individuals nested in 2186 mother-father dyads.

Formula (Wilkinson notation): children ~ s(paternalage) + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

2.4.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd
student_t(3, 0, 10)	sds

2.4.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.47 [0.38;0.56]	0.22 [0.20;0.25]

2.4.3.1 Splines

Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
sds(spaternalage_1)	0.65 [0.02;2.32]	0.26 [0.01;0.93]

2.4.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1760-1765	1.00 [0.88;1.12]	0.95 [0.67;1.33]
birth cohort 1765-1770	0.89 [0.79;0.99]	0.73 [0.55;0.98]
birth cohort 1770-1775	0.89 [0.80;1.00]	0.94 [0.70;1.25]
birth cohort 1775-1780	0.97 [0.88;1.09]	0.82 [0.61;1.08]
birth cohort 1780-1785	0.89 [0.80;1.00]	0.76 [0.57;1.02]
birth cohort 1785-1790	0.91 [0.82;1.01]	0.66 [0.49;0.87]
birth cohort 1790-1795	0.93 [0.84;1.03]	0.73 [0.55;0.96]
birth cohort 1795-1800	0.90 [0.82;0.99]	0.63 [0.49;0.81]
birth cohort 1800-1805	0.89 [0.81;0.98]	0.58 [0.45;0.74]
birth cohort 1805-1810	0.87 [0.79;0.96]	0.76 [0.60;0.98]
birth cohort 1810-1815	0.91 [0.83;0.99]	0.64 [0.50;0.82]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1815-1820	0.87 [0.79;0.94]	0.49 [0.39;0.62]
birth cohort 1820-1825	0.83 [0.76;0.91]	0.59 [0.46;0.74]
birth cohort 1825-1830	0.81 [0.74;0.88]	0.58 [0.46;0.72]
birth cohort 1830-1835	0.83 [0.76;0.91]	0.58 [0.45;0.73]
Intercept	6.45 [4.37;9.53]	1.74 [0.58;4.99]
last born	0.96 [0.92;1.00]	1.08 [0.97;1.22]
male	1.08 [1.05;1.12]	1.32 [1.20;1.44]
maternal loss 0-1	1.10 [0.95;1.29]	4.83 [3.38;7.03]
maternal loss 1-5	0.98 [0.89;1.07]	1.79 [1.42;2.27]
maternal loss 10-15	1.03 [0.95;1.11]	1.59 [1.29;1.96]
maternal loss 15-20	1.00 [0.92;1.08]	1.35 [1.10;1.66]
maternal loss 20-25	1.00 [0.93;1.08]	1.29 [1.05;1.58]
maternal loss 25-30	0.98 [0.91;1.05]	1.20 [1.00;1.43]
maternal loss 30-35	0.95 [0.89;1.01]	1.24 [1.05;1.48]
maternal loss 35-40	0.97 [0.91;1.02]	1.06 [0.91;1.25]
maternal loss 40-45	0.97 [0.91;1.03]	1.31 [1.12;1.54]
maternal loss 5-10	1.07 [0.99;1.16]	1.61 [1.31;1.98]
maternalage factor 14-20	0.95 [0.79;1.13]	1.26 [0.80;1.97]
maternalage factor 35-50	1.00 [0.94;1.05]	1.14 [0.99;1.31]
nr siblings	1.01 [0.99;1.02]	1.11 [1.07;1.15]
older siblings 1	1.03 [0.98;1.08]	1.00 [0.86;1.16]
older siblings 2	0.96 [0.89;1.03]	0.88 [0.72;1.06]
older siblings 3	0.94 [0.86;1.03]	0.85 [0.67;1.09]
older siblings 4	0.92 [0.82;1.03]	0.82 [0.60;1.11]
older siblings 5+	0.92 [0.79;1.07]	0.58 [0.39;0.86]
paternal loss 0-1	0.86 [0.74;1.00]	1.76 [1.24;2.54]
paternal loss 1-5	0.96 [0.87;1.06]	1.69 [1.31;2.19]
paternal loss 10-15	1.01 [0.93;1.08]	1.17 [0.95;1.44]
paternal loss 15-20	0.91 [0.84;0.98]	1.10 [0.90;1.35]
paternal loss 20-25	0.89 [0.82;0.95]	1.16 [0.96;1.41]
paternal loss 25-30	0.99 [0.93;1.06]	1.06 [0.88;1.27]
paternal loss 30-35	0.97 [0.91;1.04]	0.98 [0.82;1.17]
paternal loss 35-40	0.99 [0.93;1.06]	0.99 [0.83;1.18]
paternal loss 40-45	0.99 [0.92;1.06]	1.16 [0.95;1.40]
paternal loss 5-10	0.93 [0.86;1.01]	1.21 [0.97;1.50]
paternalage mean	0.94 [0.85;1.04]	0.84 [0.64;1.12]
spaternalage	1.02 [0.88;1.14]	1.20 [0.87;1.66]

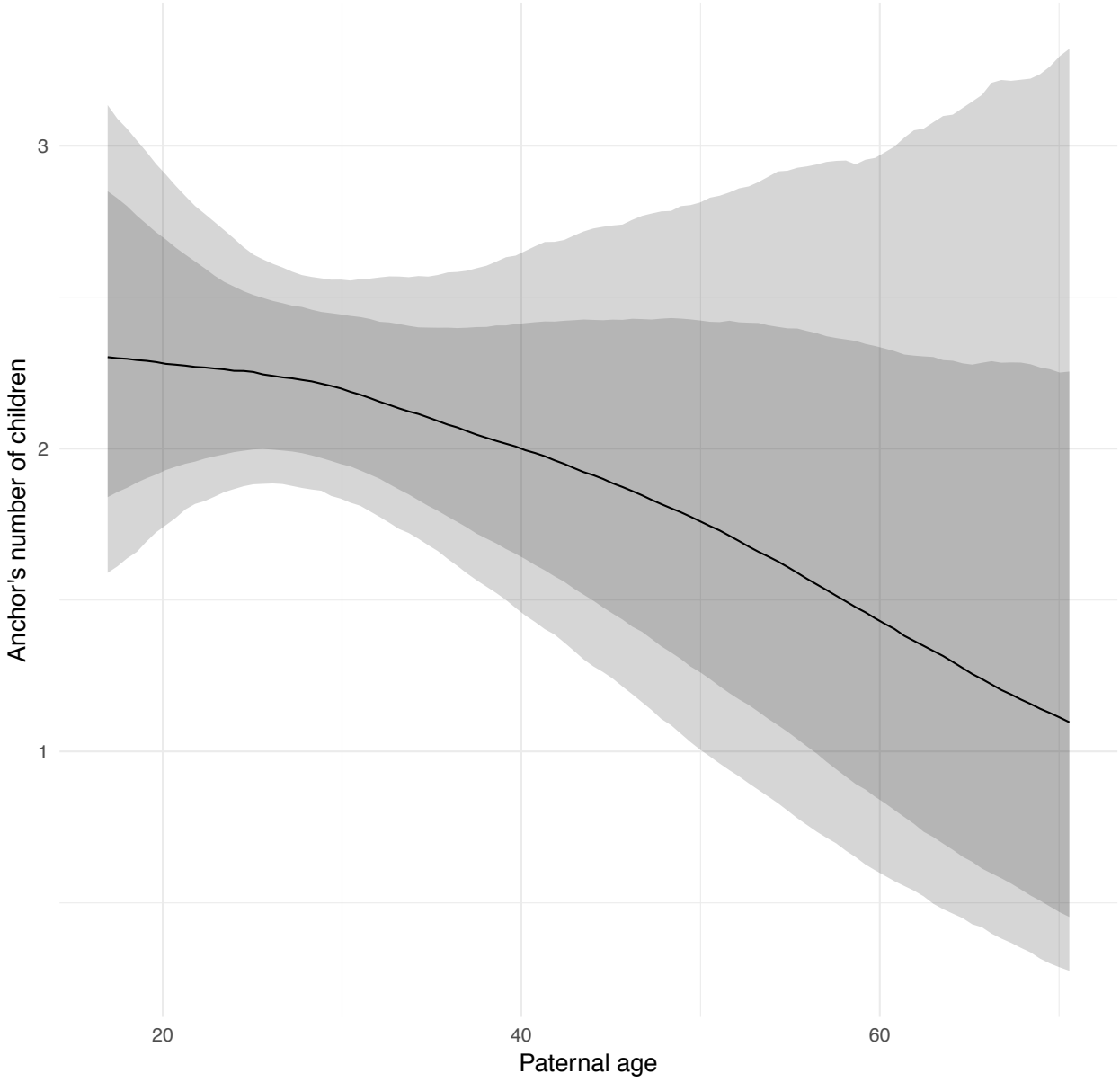
2.4.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-6.74	[-25.17; 14.85]	[-19.34; 6.80]

2.4.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



2.5 Main model comparison

We compare the four models using an approximate leave-one-out cross-validation information criterion as implemented in `brms` and `loo` and the Watanabe-Akaike information criterion.

2.5.1 Approximate leave-one-out (LOO) cross-validation

	LOOIC	SE
m1	30344	288.3
m2	30240	285
m3	30237	285.1
m4	30243	285.1
m1 - m2	104.1	40.85
m1 - m3	107.1	40.81
m1 - m4	101.2	40.9
m2 - m3	3.03	5.65
m2 - m4	-2.88	6.18
m3 - m4	-5.91	3.86

2.5.2 Watanabe-Akaike information criterion

	WAIC	SE
m1	30344	288.3
m2	30201	284.2
m3	30198	284.2
m4	30204	284.2
m1 - m2	143.2	41.13
m1 - m3	145.9	41.11
m1 - m4	140.3	41.21
m2 - m3	2.72	5.32
m2 - m4	-2.85	5.86
m3 - m4	-5.57	3.47

2.6 *e1*: Selective episode: offspring survival of the first year

In the first selective episode model, we tested how much of the paternal age effect happens in the first selective episode, i.e. in the offspring's survival of the first year.

2.6.1 Model summary

Data: 9447 individuals nested in 2186 mother-father dyads.

Formula (Wilkinson notation): survive1y ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: bernoulli
- **link**: cauchit

2.6.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

2.6.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.84 [0.53;1.10]

2.6.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1760-1765	0.96 [0.44; 2.29]
birth cohort 1765-1770	0.84 [0.39; 1.82]
birth cohort 1770-1775	1.02 [0.46; 2.29]
birth cohort 1775-1780	0.63 [0.33; 1.18]
birth cohort 1780-1785	0.67 [0.33; 1.30]
birth cohort 1785-1790	0.86 [0.41; 1.98]
birth cohort 1790-1795	1.72 [0.75; 4.36]
birth cohort 1795-1800	0.68 [0.36; 1.26]
birth cohort 1800-1805	1.46 [0.69; 3.26]
birth cohort 1805-1810	0.60 [0.32; 1.08]
birth cohort 1810-1815	0.90 [0.48; 1.68]
birth cohort 1815-1820	2.01 [0.93; 4.60]
birth cohort 1820-1825	2.42 [1.06; 6.23]
birth cohort 1825-1830	2.32 [1.05; 5.39]
birth cohort 1830-1835	1.38 [0.70; 2.83]
Intercept	56.94 [16.78;189.11]
last born	0.94 [0.65; 1.40]
male	0.71 [0.54; 0.92]
maternal loss 0-1	0.08 [0.04; 0.14]

Effect	Hurdle Odds ratio
maternal loss 1-5	0.49 [0.25; 1.00]
maternal loss 10-15	0.76 [0.39; 1.55]
maternal loss 15-20	0.86 [0.45; 1.84]
maternal loss 20-25	0.82 [0.45; 1.54]
maternal loss 25-30	1.12 [0.62; 2.19]
maternal loss 30-35	0.67 [0.41; 1.13]
maternal loss 35-40	0.72 [0.44; 1.21]
maternal loss 40-45	0.78 [0.46; 1.35]
maternal loss 5-10	0.62 [0.32; 1.23]
maternalage factor 14-20	1.02 [0.39; 3.59]
maternalage factor 35-50	0.75 [0.48; 1.17]
nr siblings	0.72 [0.66; 0.78]
older siblings 1	1.66 [1.10; 2.53]
older siblings 2	2.26 [1.36; 3.82]
older siblings 3	3.37 [1.81; 6.43]
older siblings 4	3.79 [1.87; 8.22]
older siblings 5+	10.79 [3.96; 31.27]
paternal loss 0-1	0.34 [0.14; 0.94]
paternal loss 1-5	0.41 [0.21; 0.81]
paternal loss 10-15	0.63 [0.34; 1.19]
paternal loss 15-20	0.51 [0.29; 0.89]
paternal loss 20-25	0.74 [0.40; 1.31]
paternal loss 25-30	0.73 [0.40; 1.29]
paternal loss 30-35	0.91 [0.52; 1.65]
paternal loss 35-40	0.93 [0.52; 1.71]
paternal loss 40-45	0.46 [0.27; 0.79]
paternal loss 5-10	0.77 [0.37; 1.65]
paternalage	0.46 [0.23; 0.92]
paternalage mean	2.50 [1.19; 5.14]

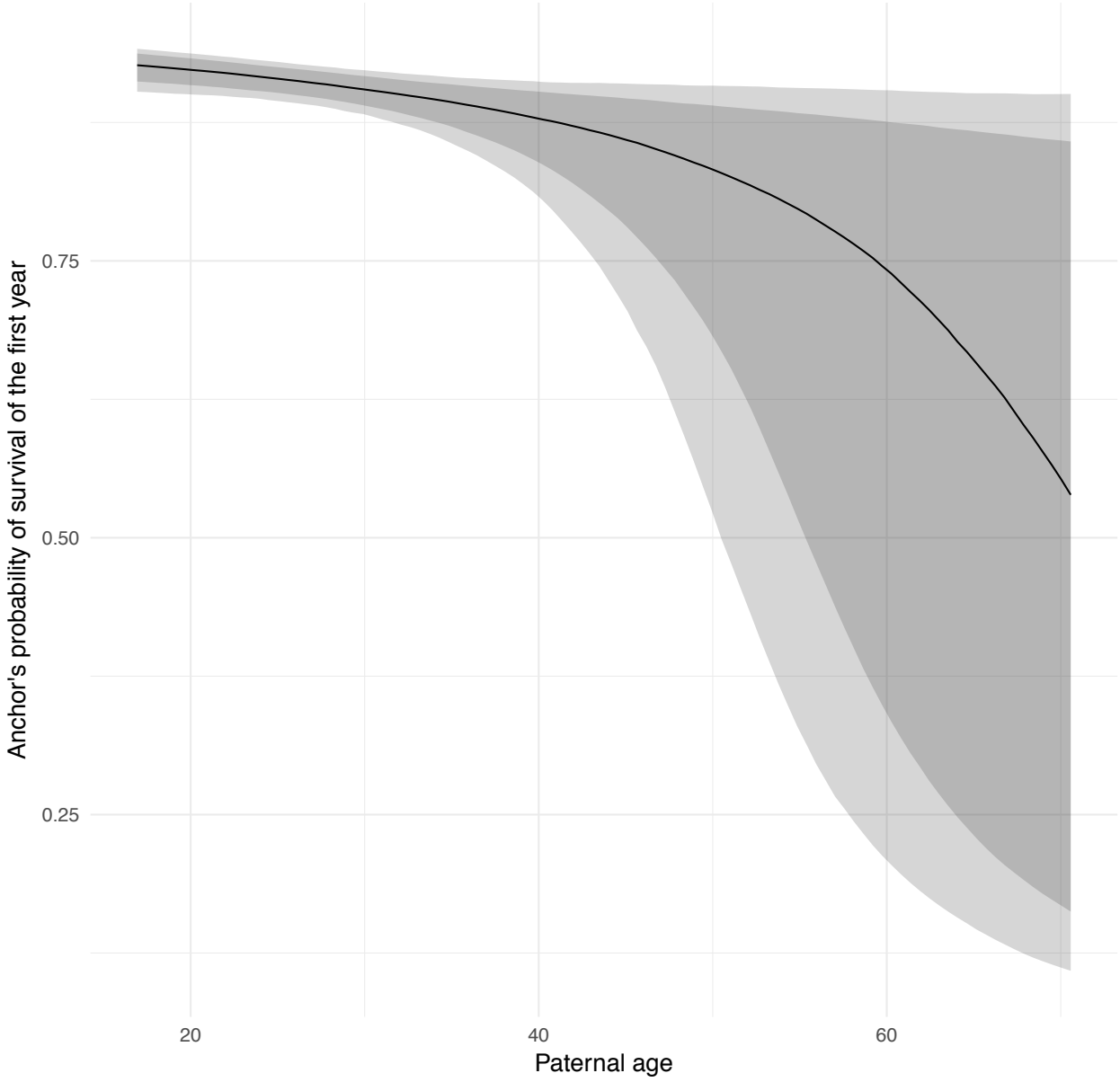
2.6.5 Paternal age effect

This is the effect of 10 years of paternal age within families on probability of survival of the first year, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-2.15	[-5.35;-0.21]	[-4.03;-0.82]

2.6.5.1 Marginal effect plot

Paternal age effect on probability of survival of the first year The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



3 Québec

3.1 *m1*: No sibling comparison

Here, we ignore the pedigree structure of the data to see whether it matters for the estimation of the paternal age effect.

3.1.1 Model summary

Data: 68724 individuals.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born.

- **family**: hurdle_poisson
- **link**: log

3.1.2 Priors

prior	class
normal(0,5)	b

3.1.3 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1675-1680	1.01 [0.98;1.04]	1.07 [0.94;1.21]
birth cohort 1680-1685	1.03 [1.01;1.06]	1.22 [1.08;1.38]
birth cohort 1685-1690	1.04 [1.01;1.07]	1.35 [1.19;1.53]
birth cohort 1690-1695	1.03 [1.01;1.06]	1.08 [0.96;1.21]
birth cohort 1695-1700	1.03 [1.00;1.05]	1.10 [0.98;1.23]
birth cohort 1700-1705	1.01 [0.99;1.04]	1.26 [1.13;1.40]
birth cohort 1705-1710	0.99 [0.97;1.02]	1.15 [1.03;1.28]
birth cohort 1710-1715	0.99 [0.96;1.01]	1.48 [1.33;1.64]
birth cohort 1715-1720	0.96 [0.93;0.98]	1.32 [1.19;1.47]
birth cohort 1720-1725	0.96 [0.94;0.98]	1.35 [1.22;1.50]
birth cohort 1725-1730	0.94 [0.91;0.96]	1.89 [1.70;2.09]
birth cohort 1730-1735	0.96 [0.93;0.98]	1.98 [1.80;2.18]
birth cohort 1735-1740	0.95 [0.92;0.97]	1.70 [1.54;1.87]
Intercept	8.47 [8.20;8.75]	0.42 [0.37;0.48]
last born	1.01 [0.99;1.02]	1.02 [0.97;1.08]
male	1.12 [1.11;1.13]	1.49 [1.45;1.54]
maternal loss 0-1	0.97 [0.93;1.01]	2.84 [2.50;3.24]
maternal loss 1-5	0.98 [0.96;1.01]	1.52 [1.41;1.65]
maternal loss 10-15	0.99 [0.97;1.01]	1.30 [1.21;1.40]
maternal loss 15-20	1.00 [0.98;1.02]	1.22 [1.13;1.30]
maternal loss 20-25	0.97 [0.96;0.99]	1.18 [1.10;1.26]
maternal loss 25-30	0.98 [0.96;0.99]	1.07 [1.00;1.14]
maternal loss 30-35	0.98 [0.97;1.00]	1.09 [1.03;1.15]
maternal loss 35-40	0.99 [0.98;1.00]	1.08 [1.02;1.15]
maternal loss 40-45	1.00 [0.98;1.01]	1.00 [0.95;1.06]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
maternal loss 5-10	1.01 [0.99;1.02]	1.34 [1.25;1.44]
maternal loss unclear	0.98 [0.96;0.99]	1.21 [1.15;1.28]
maternalage factor 14-20	0.99 [0.98;1.01]	1.00 [0.93;1.06]
maternalage factor 35-50	1.01 [1.00;1.02]	1.07 [1.02;1.12]
nr siblings	1.01 [1.01;1.01]	1.02 [1.01;1.02]
older siblings 1	0.98 [0.97;1.00]	0.95 [0.90;1.01]
older siblings 2	0.99 [0.97;1.00]	0.93 [0.87;0.99]
older siblings 3	0.98 [0.96;0.99]	0.89 [0.84;0.95]
older siblings 4	1.00 [0.98;1.01]	0.89 [0.83;0.95]
older siblings 5+	0.99 [0.97;1.00]	0.86 [0.81;0.92]
paternal loss 0-1	0.99 [0.96;1.03]	1.81 [1.59;2.06]
paternal loss 1-5	1.00 [0.98;1.02]	1.40 [1.27;1.53]
paternal loss 10-15	0.99 [0.97;1.00]	1.21 [1.12;1.31]
paternal loss 15-20	0.98 [0.96;0.99]	1.32 [1.23;1.42]
paternal loss 20-25	0.98 [0.96;0.99]	1.20 [1.12;1.28]
paternal loss 25-30	0.99 [0.97;1.00]	1.20 [1.13;1.28]
paternal loss 30-35	0.98 [0.96;0.99]	1.15 [1.08;1.22]
paternal loss 35-40	0.98 [0.96;0.99]	1.11 [1.04;1.18]
paternal loss 40-45	1.00 [0.98;1.01]	1.06 [0.99;1.14]
paternal loss 5-10	1.00 [0.98;1.02]	1.37 [1.27;1.49]
paternal loss unclear	0.96 [0.95;0.98]	1.42 [1.34;1.51]
paternalage	0.98 [0.98;0.99]	1.00 [0.97;1.02]

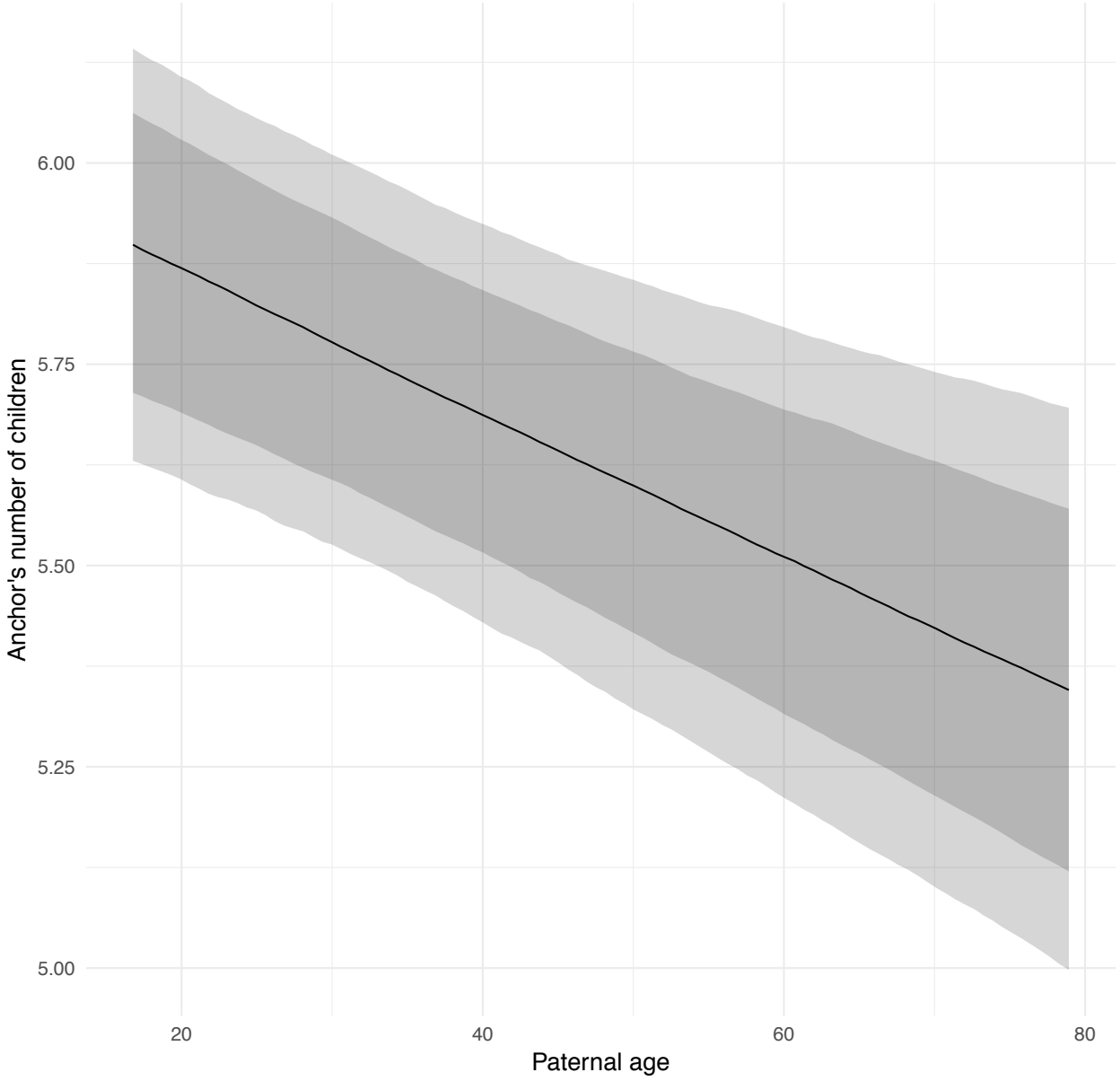
3.1.4 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-1.57	[-2.51;-0.59]	[-2.17;-0.90]

3.1.4.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



3.2 *m2*: Sibling comparison, no paternal age effect

Here, we compared siblings by including a random intercept for the family, but we modelled no effect for paternal age differences among siblings.

3.2.1 Model summary

Data: 68724 individuals nested in 12205 mother-father dyads.

Formula (Wilkinson notation): children ~ birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

3.2.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

3.2.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.63 [0.60;0.66]	0.27 [0.27;0.28]

3.2.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1675-1680	1.02 [0.99;1.05]	1.08 [0.94;1.26]
birth cohort 1680-1685	1.05 [1.02;1.08]	1.26 [1.10;1.45]
birth cohort 1685-1690	1.05 [1.02;1.09]	1.41 [1.23;1.62]
birth cohort 1690-1695	1.05 [1.02;1.08]	1.09 [0.96;1.26]
birth cohort 1695-1700	1.04 [1.00;1.07]	1.11 [0.98;1.27]
birth cohort 1700-1705	1.02 [0.99;1.06]	1.30 [1.14;1.48]
birth cohort 1705-1710	1.00 [0.97;1.03]	1.17 [1.03;1.33]
birth cohort 1710-1715	0.99 [0.96;1.02]	1.55 [1.37;1.77]
birth cohort 1715-1720	0.96 [0.93;1.00]	1.34 [1.19;1.53]
birth cohort 1720-1725	0.96 [0.93;0.99]	1.37 [1.21;1.55]
birth cohort 1725-1730	0.94 [0.91;0.97]	1.97 [1.75;2.24]
birth cohort 1730-1735	0.95 [0.93;0.98]	2.08 [1.85;2.35]
birth cohort 1735-1740	0.94 [0.92;0.97]	1.77 [1.58;2.00]
Intercept	8.26 [7.82;8.71]	0.43 [0.36;0.51]
last born	1.00 [0.99;1.02]	1.01 [0.95;1.07]
male	1.12 [1.11;1.13]	1.55 [1.50;1.60]
maternal loss 0-1	0.95 [0.91;1.00]	2.99 [2.59;3.45]
maternal loss 1-5	0.97 [0.95;1.00]	1.52 [1.38;1.67]
maternal loss 10-15	0.99 [0.97;1.02]	1.31 [1.21;1.43]
maternal loss 15-20	1.01 [0.98;1.03]	1.23 [1.14;1.34]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
maternal loss 20-25	0.98 [0.96;1.00]	1.20 [1.11;1.30]
maternal loss 25-30	0.99 [0.97;1.01]	1.07 [0.99;1.15]
maternal loss 30-35	0.99 [0.97;1.01]	1.10 [1.03;1.18]
maternal loss 35-40	1.00 [0.98;1.02]	1.09 [1.02;1.16]
maternal loss 40-45	1.00 [0.99;1.02]	0.99 [0.93;1.07]
maternal loss 5-10	1.01 [0.99;1.04]	1.32 [1.21;1.44]
maternal loss unclear	0.98 [0.96;1.01]	1.26 [1.17;1.36]
maternalage factor 14-20	0.99 [0.97;1.01]	1.04 [0.97;1.11]
maternalage factor 35-50	1.01 [0.99;1.02]	1.09 [1.04;1.14]
nr siblings	1.01 [1.00;1.01]	1.01 [1.01;1.02]
older siblings 1	0.98 [0.97;0.99]	0.96 [0.90;1.02]
older siblings 2	0.98 [0.96;0.99]	0.93 [0.87;1.00]
older siblings 3	0.97 [0.96;0.99]	0.90 [0.84;0.96]
older siblings 4	0.98 [0.97;1.00]	0.90 [0.84;0.97]
older siblings 5+	0.97 [0.96;0.99]	0.87 [0.82;0.93]
paternal loss 0-1	0.97 [0.93;1.01]	1.85 [1.61;2.13]
paternal loss 1-5	0.99 [0.96;1.02]	1.41 [1.27;1.57]
paternal loss 10-15	0.99 [0.96;1.01]	1.22 [1.12;1.33]
paternal loss 15-20	0.98 [0.96;1.00]	1.35 [1.24;1.47]
paternal loss 20-25	0.98 [0.96;1.00]	1.21 [1.12;1.30]
paternal loss 25-30	0.99 [0.97;1.01]	1.21 [1.12;1.31]
paternal loss 30-35	0.98 [0.96;1.00]	1.16 [1.08;1.25]
paternal loss 35-40	0.98 [0.96;1.00]	1.11 [1.03;1.19]
paternal loss 40-45	1.00 [0.98;1.02]	1.07 [0.99;1.16]
paternal loss 5-10	0.99 [0.97;1.02]	1.38 [1.26;1.52]
paternal loss unclear	0.95 [0.93;0.97]	1.47 [1.37;1.59]
paternalage mean	0.98 [0.97;0.99]	0.98 [0.95;1.01]

3.2.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

This model did not contain a within family paternal age predictor.

3.3 *m3*: Sibling comparison, linear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a linear effect for paternal age differences among siblings.

3.3.1 Model summary

Data: 68724 individuals nested in 12205 mother-father dyads.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

3.3.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

3.3.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.63 [0.60;0.66]	0.27 [0.27;0.28]

3.3.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1675-1680	1.02 [0.99;1.05]	1.08 [0.95;1.24]
birth cohort 1680-1685	1.05 [1.02;1.08]	1.26 [1.10;1.44]
birth cohort 1685-1690	1.05 [1.02;1.09]	1.40 [1.22;1.60]
birth cohort 1690-1695	1.05 [1.02;1.08]	1.08 [0.95;1.24]
birth cohort 1695-1700	1.04 [1.00;1.07]	1.10 [0.98;1.26]
birth cohort 1700-1705	1.02 [0.99;1.05]	1.29 [1.14;1.46]
birth cohort 1705-1710	1.00 [0.97;1.03]	1.17 [1.03;1.32]
birth cohort 1710-1715	0.99 [0.96;1.02]	1.54 [1.37;1.75]
birth cohort 1715-1720	0.96 [0.93;0.99]	1.33 [1.18;1.50]
birth cohort 1720-1725	0.96 [0.93;0.99]	1.36 [1.21;1.54]
birth cohort 1725-1730	0.94 [0.91;0.96]	1.95 [1.73;2.21]
birth cohort 1730-1735	0.95 [0.92;0.98]	2.06 [1.84;2.33]
birth cohort 1735-1740	0.94 [0.91;0.97]	1.76 [1.57;1.98]
Intercept	8.30 [7.86;8.77]	0.45 [0.38;0.53]
last born	1.00 [0.99;1.02]	1.00 [0.94;1.05]
male	1.12 [1.11;1.13]	1.55 [1.50;1.60]
maternal loss 0-1	0.95 [0.91;0.99]	2.88 [2.51;3.35]
maternal loss 1-5	0.97 [0.94;1.00]	1.48 [1.35;1.64]
maternal loss 10-15	0.99 [0.97;1.02]	1.29 [1.19;1.40]
maternal loss 15-20	1.01 [0.98;1.03]	1.22 [1.13;1.33]
maternal loss 20-25	0.98 [0.96;1.00]	1.19 [1.09;1.29]
maternal loss 25-30	0.99 [0.97;1.01]	1.05 [0.98;1.14]
maternal loss 30-35	0.99 [0.97;1.01]	1.09 [1.01;1.17]
maternal loss 35-40	1.00 [0.98;1.02]	1.08 [1.02;1.16]
maternal loss 40-45	1.00 [0.99;1.02]	0.99 [0.92;1.06]
maternal loss 5-10	1.01 [0.99;1.04]	1.29 [1.19;1.41]
maternal loss unclear	0.98 [0.96;1.01]	1.25 [1.16;1.34]
maternalage factor 14-20	0.99 [0.97;1.01]	1.05 [0.97;1.13]
maternalage factor 35-50	1.00 [0.99;1.02]	1.03 [0.97;1.09]
nr siblings	1.01 [1.00;1.01]	1.02 [1.01;1.03]
older siblings 1	0.98 [0.96;0.99]	0.94 [0.89;1.00]
older siblings 2	0.97 [0.96;0.99]	0.90 [0.84;0.96]
older siblings 3	0.97 [0.95;0.99]	0.85 [0.79;0.91]
older siblings 4	0.98 [0.96;1.00]	0.83 [0.76;0.90]
older siblings 5+	0.96 [0.94;0.99]	0.76 [0.69;0.85]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
paternal loss 0-1	0.96 [0.92;1.01]	1.78 [1.54;2.05]
paternal loss 1-5	0.98 [0.95;1.02]	1.37 [1.24;1.52]
paternal loss 10-15	0.98 [0.96;1.01]	1.19 [1.09;1.29]
paternal loss 15-20	0.98 [0.95;1.00]	1.33 [1.22;1.45]
paternal loss 20-25	0.98 [0.96;1.00]	1.19 [1.10;1.29]
paternal loss 25-30	0.99 [0.97;1.01]	1.20 [1.11;1.29]
paternal loss 30-35	0.98 [0.96;1.00]	1.15 [1.07;1.24]
paternal loss 35-40	0.98 [0.96;1.00]	1.11 [1.03;1.19]
paternal loss 40-45	1.00 [0.98;1.02]	1.07 [0.99;1.16]
paternal loss 5-10	0.99 [0.96;1.02]	1.34 [1.22;1.47]
paternal loss unclear	0.95 [0.92;0.97]	1.45 [1.34;1.57]
paternalage	1.01 [0.99;1.03]	1.14 [1.05;1.23]
paternalage mean	0.97 [0.95;0.99]	0.87 [0.80;0.94]

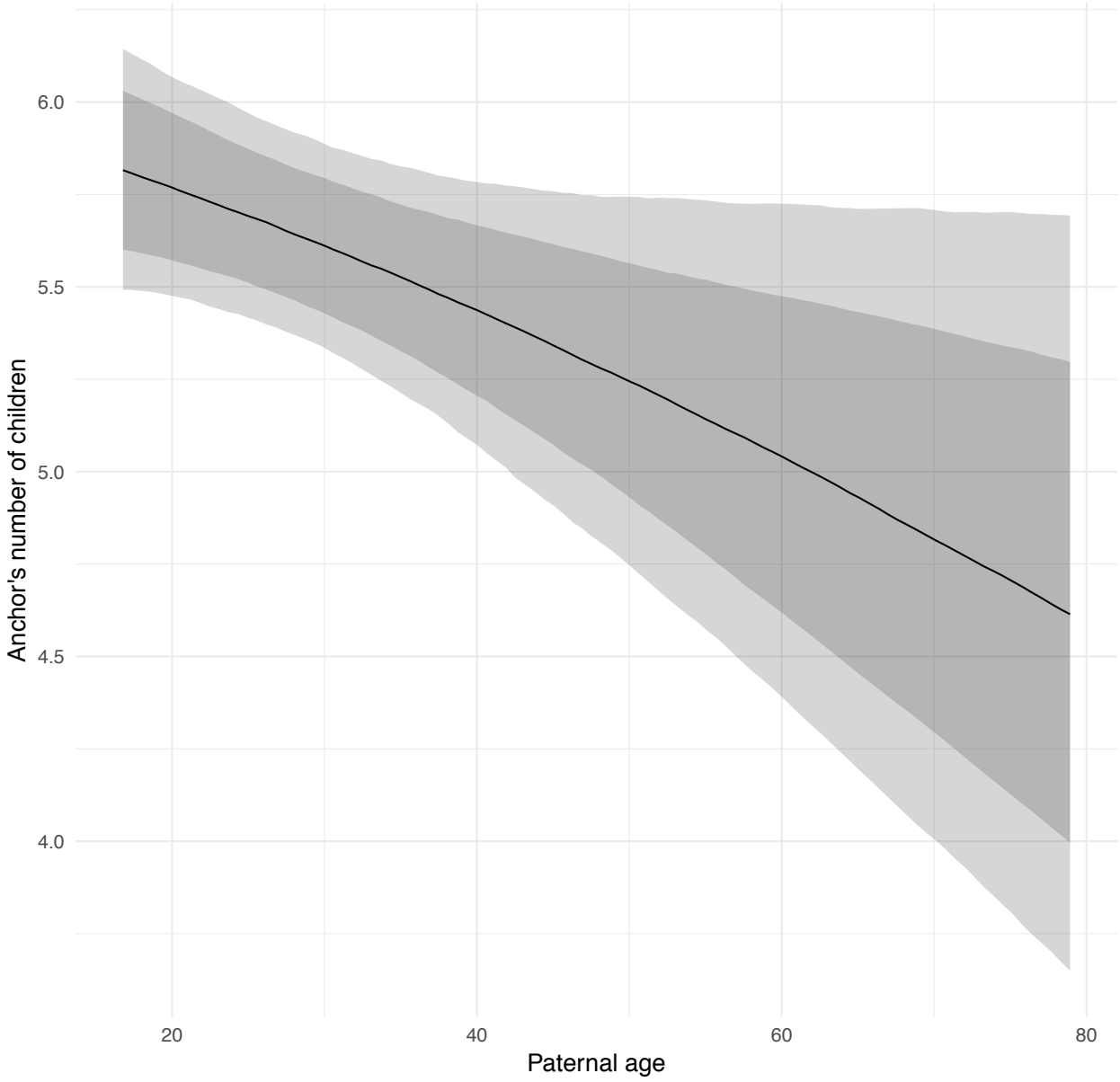
3.3.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-3.00	[-6.08; 0.24]	[-4.97;-0.90]

3.3.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



3.4 *m4*: Sibling comparison, nonlinear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a possibly nonlinear effect for paternal age differences among siblings.

3.4.1 Model summary

Data: 68724 individuals nested in 12205 mother-father dyads.

Formula (Wilkinson notation): children ~ s(paternalage) + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

3.4.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd
student_t(3, 0, 10)	sds

3.4.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.63 [0.60;0.65]	0.27 [0.27;0.28]

3.4.3.1 Splines

Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
sds(spaternalage_1)	0.80 [0.24;1.97]	0.10 [0.00;0.32]

3.4.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1675-1680	1.02 [0.99;1.05]	1.08 [0.95;1.23]
birth cohort 1680-1685	1.05 [1.02;1.08]	1.26 [1.10;1.44]
birth cohort 1685-1690	1.05 [1.02;1.09]	1.41 [1.23;1.62]
birth cohort 1690-1695	1.05 [1.02;1.08]	1.10 [0.97;1.26]
birth cohort 1695-1700	1.04 [1.00;1.07]	1.12 [0.99;1.28]
birth cohort 1700-1705	1.02 [0.99;1.05]	1.31 [1.16;1.48]
birth cohort 1705-1710	1.00 [0.97;1.03]	1.18 [1.04;1.34]
birth cohort 1710-1715	0.99 [0.96;1.02]	1.56 [1.38;1.76]
birth cohort 1715-1720	0.96 [0.93;0.99]	1.35 [1.20;1.52]
birth cohort 1720-1725	0.96 [0.93;0.99]	1.38 [1.23;1.55]
birth cohort 1725-1730	0.94 [0.91;0.96]	1.98 [1.77;2.23]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1730-1735	0.95 [0.92;0.98]	2.09 [1.87;2.35]
birth cohort 1735-1740	0.94 [0.91;0.97]	1.78 [1.58;2.00]
Intercept	8.63 [7.82;9.56]	0.82 [0.58;1.18]
last born	1.00 [0.99;1.02]	1.00 [0.94;1.06]
male	1.12 [1.11;1.13]	1.55 [1.50;1.60]
maternal loss 0-1	0.95 [0.91;0.99]	2.89 [2.54;3.32]
maternal loss 1-5	0.97 [0.94;1.00]	1.48 [1.35;1.64]
maternal loss 10-15	0.99 [0.97;1.02]	1.30 [1.19;1.41]
maternal loss 15-20	1.01 [0.98;1.03]	1.22 [1.13;1.34]
maternal loss 20-25	0.98 [0.96;1.00]	1.19 [1.10;1.28]
maternal loss 25-30	0.99 [0.97;1.01]	1.05 [0.98;1.14]
maternal loss 30-35	0.99 [0.97;1.01]	1.09 [1.01;1.17]
maternal loss 35-40	1.00 [0.98;1.02]	1.08 [1.02;1.15]
maternal loss 40-45	1.00 [0.99;1.02]	0.99 [0.93;1.05]
maternal loss 5-10	1.01 [0.99;1.04]	1.30 [1.19;1.41]
maternal loss unclear	0.98 [0.96;1.01]	1.25 [1.16;1.35]
maternalage factor 14-20	0.99 [0.97;1.01]	1.06 [0.98;1.14]
maternalage factor 35-50	1.00 [0.99;1.02]	1.04 [0.98;1.10]
nr siblings	1.01 [1.00;1.01]	1.02 [1.01;1.03]
older siblings 1	0.98 [0.96;0.99]	0.92 [0.87;0.98]
older siblings 2	0.98 [0.96;0.99]	0.86 [0.81;0.93]
older siblings 3	0.97 [0.95;0.99]	0.80 [0.74;0.87]
older siblings 4	0.98 [0.96;1.00]	0.77 [0.71;0.84]
older siblings 5+	0.96 [0.94;0.99]	0.70 [0.63;0.78]
paternal loss 0-1	0.96 [0.92;1.00]	1.78 [1.56;2.05]
paternal loss 1-5	0.98 [0.95;1.01]	1.37 [1.23;1.52]
paternal loss 10-15	0.98 [0.96;1.01]	1.18 [1.08;1.29]
paternal loss 15-20	0.98 [0.95;1.00]	1.32 [1.21;1.43]
paternal loss 20-25	0.98 [0.96;1.00]	1.18 [1.09;1.27]
paternal loss 25-30	0.99 [0.97;1.01]	1.18 [1.09;1.27]
paternal loss 30-35	0.98 [0.96;1.00]	1.13 [1.06;1.22]
paternal loss 35-40	0.98 [0.96;1.00]	1.09 [1.02;1.17]
paternal loss 40-45	1.00 [0.98;1.02]	1.06 [0.98;1.14]
paternal loss 5-10	0.99 [0.96;1.02]	1.34 [1.22;1.47]
paternal loss unclear	0.95 [0.92;0.97]	1.44 [1.34;1.56]
paternalage mean	0.97 [0.95;0.99]	0.84 [0.77;0.91]
spaternalage	1.02 [0.99;1.10]	1.12 [0.89;1.41]

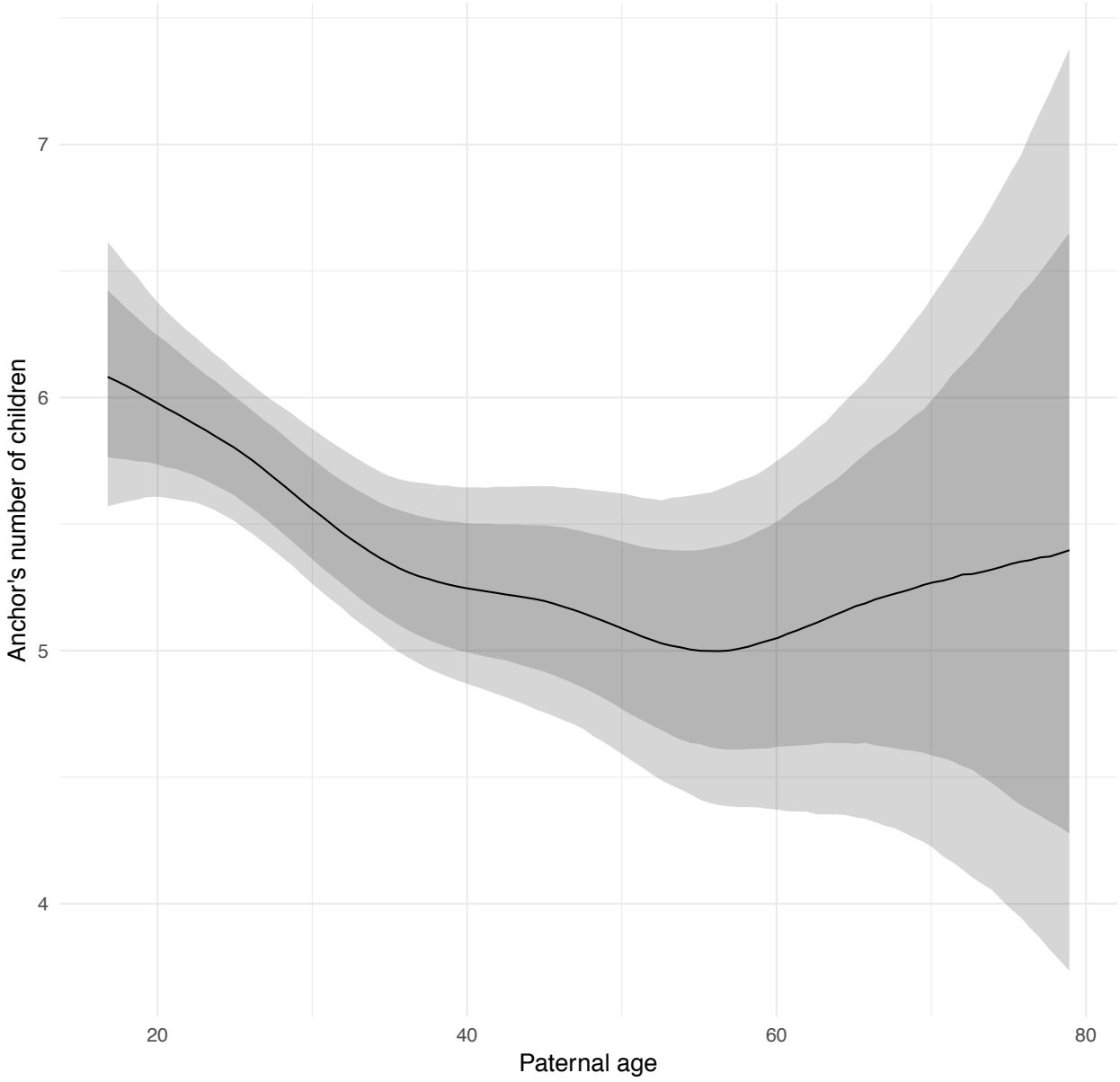
3.4.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-7.98	[-11.85; -4.12]	[-10.49; -5.23]

3.4.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



3.5 Main model comparison

We compare the four models using an approximate leave-one-out cross-validation information criterion as implemented in `brms` and `loo` and the Watanabe-Akaike information criterion.

3.5.1 Approximate leave-one-out (LOO) cross-validation

	LOOIC	SE
m1	299887	992.2
m2	293615	957.5
m3	293619	957.7
m4	293606	957.7
m1 - m2	6272	299.1
m1 - m3	6268	299.3
m1 - m4	6281	299.3
m2 - m3	-3.66	17.39
m2 - m4	9.83	19.87
m3 - m4	13.49	18.8

3.5.2 Watanabe-Akaike information criterion

	WAIC	SE
m1	299887	992.2
m2	292565	950.5
m3	292572	950.7
m4	292556	950.7
m1 - m2	7322	299.6
m1 - m3	7315	299.9
m1 - m4	7331	299.9
m2 - m3	-7.05	13.1
m2 - m4	8.81	16.28
m3 - m4	15.86	14.79

3.6 *e1*: Selective episode: offspring survival of the first year

In the first selective episode model, we tested how much of the paternal age effect happens in the first selective episode, i.e. in the offspring's survival of the first year.

3.6.1 Model summary

Data: 61493 individuals nested in 11940 mother-father dyads.

Formula (Wilkinson notation): survive1y ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: bernoulli
- **link**: cauchit

3.6.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

3.6.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	1.06 [1.00;1.12]

3.6.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1675-1680	1.22 [0.69; 2.17]
birth cohort 1680-1685	0.57 [0.34; 0.92]
birth cohort 1685-1690	0.23 [0.15; 0.35]
birth cohort 1690-1695	0.43 [0.28; 0.67]
birth cohort 1695-1700	0.44 [0.28; 0.66]
birth cohort 1700-1705	0.27 [0.17; 0.41]
birth cohort 1705-1710	0.46 [0.30; 0.70]
birth cohort 1710-1715	0.24 [0.15; 0.36]
birth cohort 1715-1720	0.28 [0.18; 0.41]
birth cohort 1720-1725	0.28 [0.18; 0.41]
birth cohort 1725-1730	0.19 [0.13; 0.28]
birth cohort 1730-1735	0.15 [0.10; 0.23]
birth cohort 1735-1740	0.19 [0.12; 0.28]
Intercept	42.30 [25.95;69.83]
last born	1.00 [0.90; 1.10]
male	0.66 [0.62; 0.70]
maternal loss 0-1	0.11 [0.09; 0.13]
maternal loss 1-5	0.50 [0.42; 0.59]
maternal loss 10-15	0.67 [0.57; 0.78]

Effect	Hurdle Odds ratio
maternal loss 15-20	0.68 [0.59; 0.79]
maternal loss 20-25	0.76 [0.66; 0.89]
maternal loss 25-30	0.85 [0.73; 0.97]
maternal loss 30-35	0.81 [0.71; 0.92]
maternal loss 35-40	0.79 [0.70; 0.89]
maternal loss 40-45	0.97 [0.85; 1.11]
maternal loss 5-10	0.64 [0.55; 0.76]
maternal loss unclear	0.76 [0.66; 0.88]
maternalage factor 14-20	0.84 [0.73; 0.96]
maternalage factor 35-50	0.92 [0.82; 1.02]
nr siblings	0.91 [0.90; 0.92]
older siblings 1	1.56 [1.39; 1.75]
older siblings 2	1.72 [1.52; 1.94]
older siblings 3	2.21 [1.92; 2.56]
older siblings 4	2.29 [1.96; 2.67]
older siblings 5+	2.78 [2.33; 3.32]
paternal loss 0-1	0.36 [0.29; 0.45]
paternal loss 1-5	0.62 [0.51; 0.75]
paternal loss 10-15	0.80 [0.68; 0.94]
paternal loss 15-20	0.72 [0.62; 0.84]
paternal loss 20-25	0.86 [0.74; 0.99]
paternal loss 25-30	0.78 [0.68; 0.90]
paternal loss 30-35	0.83 [0.72; 0.95]
paternal loss 35-40	0.85 [0.75; 0.97]
paternal loss 40-45	0.97 [0.83; 1.12]
paternal loss 5-10	0.70 [0.59; 0.84]
paternal loss unclear	0.70 [0.60; 0.81]
paternalage	0.64 [0.56; 0.74]
paternalage mean	1.72 [1.49; 1.99]

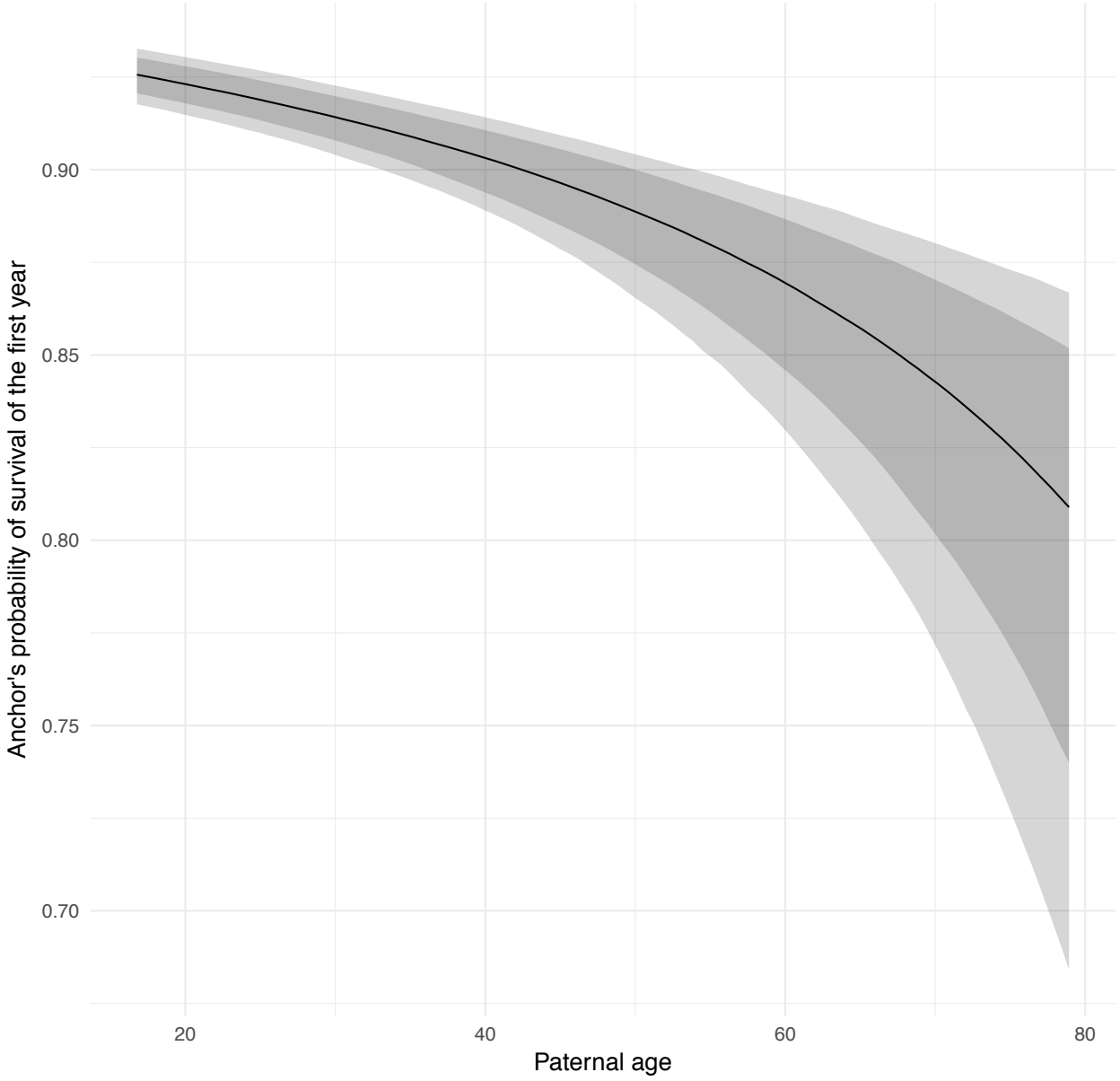
3.6.5 Paternal age effect

This is the effect of 10 years of paternal age within families on probability of survival of the first year, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-1.03	[-1.51;-0.67]	[-1.32;-0.78]

3.6.5.1 Marginal effect plot

Paternal age effect on probability of survival of the first year The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



4 Historical Sweden

4.1 *m1*: No sibling comparison

Here, we ignore the pedigree structure of the data to see whether it matters for the estimation of the paternal age effect.

4.1.1 Model summary

Data: 56663 individuals.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born.

- **family**: hurdle_poisson
- **link**: log

4.1.2 Priors

prior	class
normal(0,5)	b

4.1.3 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1750-1755	0.91 [0.75;1.10]	1.40 [0.87;2.28]
birth cohort 1755-1760	1.11 [0.96;1.29]	1.12 [0.77;1.65]
birth cohort 1760-1765	1.21 [1.07;1.37]	1.02 [0.72;1.44]
birth cohort 1765-1770	1.15 [1.01;1.30]	0.78 [0.56;1.11]
birth cohort 1770-1775	1.13 [0.99;1.28]	0.96 [0.68;1.36]
birth cohort 1775-1780	1.09 [0.97;1.24]	1.12 [0.82;1.55]
birth cohort 1780-1785	1.20 [1.06;1.36]	1.05 [0.76;1.45]
birth cohort 1785-1790	1.19 [1.06;1.34]	1.17 [0.87;1.61]
birth cohort 1790-1795	1.08 [0.97;1.22]	1.41 [1.06;1.87]
birth cohort 1795-1800	1.11 [1.00;1.24]	1.17 [0.89;1.55]
birth cohort 1800-1805	1.05 [0.94;1.18]	1.12 [0.85;1.48]
birth cohort 1805-1810	1.07 [0.96;1.20]	1.06 [0.80;1.41]
birth cohort 1810-1815	1.09 [0.97;1.21]	1.16 [0.89;1.53]
birth cohort 1815-1820	1.16 [1.04;1.29]	0.98 [0.75;1.28]
birth cohort 1820-1825	1.16 [1.05;1.29]	0.89 [0.68;1.16]
birth cohort 1825-1830	1.12 [1.01;1.24]	0.88 [0.68;1.15]
birth cohort 1830-1835	1.14 [1.03;1.27]	0.90 [0.69;1.18]
birth cohort 1835-1840	1.13 [1.02;1.26]	0.90 [0.69;1.17]
birth cohort 1840-1845	1.11 [1.00;1.23]	0.91 [0.70;1.19]
birth cohort 1845-1850	1.12 [1.01;1.24]	0.93 [0.71;1.21]
Intercept	3.83 [3.42;4.28]	0.93 [0.71;1.24]
last born	0.99 [0.97;1.00]	1.00 [0.95;1.05]
male	1.04 [1.02;1.05]	1.05 [1.01;1.08]
maternal loss 0-1	1.08 [0.99;1.18]	5.18 [4.22;6.43]
maternal loss 1-5	1.03 [0.98;1.07]	2.42 [2.17;2.72]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
maternal loss 10-15	0.95 [0.92;0.98]	2.04 [1.86;2.23]
maternal loss 15-20	0.96 [0.93;0.99]	1.79 [1.65;1.95]
maternal loss 20-25	0.93 [0.90;0.95]	1.50 [1.39;1.62]
maternal loss 25-30	0.96 [0.94;0.98]	1.33 [1.25;1.43]
maternal loss 30-35	0.97 [0.95;0.99]	1.27 [1.19;1.34]
maternal loss 35-40	0.99 [0.97;1.01]	1.16 [1.10;1.23]
maternal loss 40-45	0.98 [0.96;1.00]	1.09 [1.03;1.16]
maternal loss 5-10	0.98 [0.95;1.02]	2.12 [1.93;2.32]
maternalage factor 10-20	1.04 [0.99;1.09]	1.04 [0.90;1.21]
maternalage factor 35-59	1.07 [1.05;1.08]	1.07 [1.02;1.12]
nr siblings	1.03 [1.03;1.03]	1.03 [1.02;1.04]
older siblings 1	1.00 [0.98;1.02]	0.99 [0.94;1.05]
older siblings 2	1.00 [0.98;1.02]	1.03 [0.97;1.09]
older siblings 3	0.99 [0.97;1.01]	1.04 [0.98;1.11]
older siblings 4	0.96 [0.94;0.99]	1.03 [0.95;1.11]
older siblings 5+	0.95 [0.92;0.98]	0.98 [0.90;1.06]
paternal loss 0-1	1.09 [1.03;1.15]	2.12 [1.81;2.49]
paternal loss 1-5	1.02 [0.98;1.06]	1.81 [1.64;2.00]
paternal loss 10-15	0.97 [0.95;1.00]	1.62 [1.50;1.75]
paternal loss 15-20	0.93 [0.91;0.95]	1.47 [1.38;1.58]
paternal loss 20-25	0.97 [0.95;0.99]	1.35 [1.27;1.44]
paternal loss 25-30	0.98 [0.96;1.00]	1.24 [1.17;1.33]
paternal loss 30-35	0.98 [0.96;1.00]	1.18 [1.10;1.25]
paternal loss 35-40	1.02 [1.00;1.04]	1.13 [1.06;1.20]
paternal loss 40-45	1.02 [0.99;1.04]	1.05 [0.98;1.12]
paternal loss 5-10	0.98 [0.95;1.01]	1.87 [1.72;2.03]
paternalage	1.01 [1.00;1.01]	0.96 [0.94;0.99]

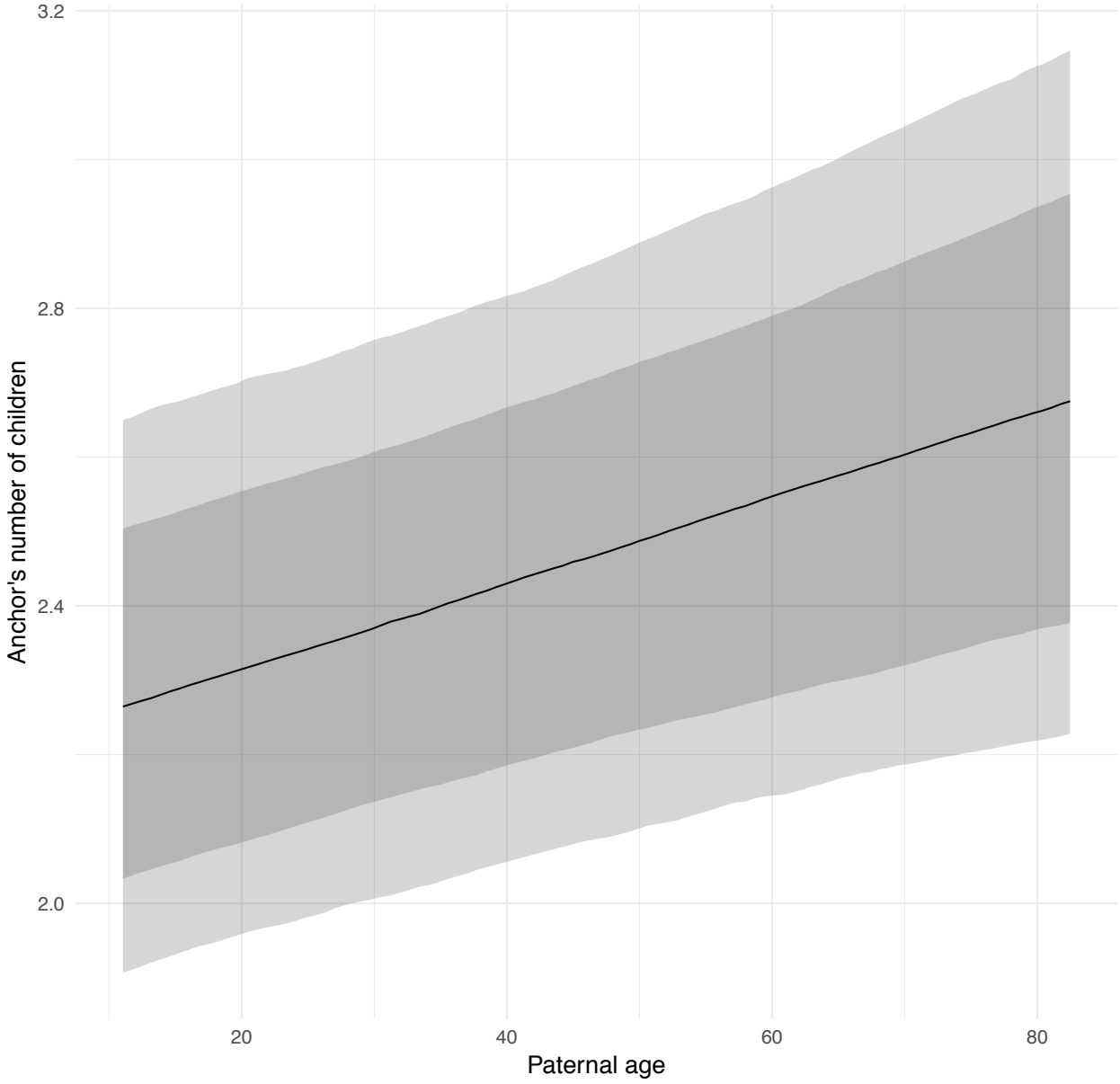
4.1.4 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	2.37	[0.69;4.08]	[1.27;3.49]

4.1.4.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



4.2 *m2*: Sibling comparison, no paternal age effect

Here, we compared siblings by including a random intercept for the family, but we modelled no effect for paternal age differences among siblings.

4.2.1 Model summary

Data: 56663 individuals nested in 14746 mother-father dyads.

Formula (Wilkinson notation): children ~ birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

4.2.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

4.2.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.82 [0.79;0.85]	0.36 [0.34;0.37]

4.2.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1750-1755	0.85 [0.67;1.07]	1.43 [0.80;2.57]
birth cohort 1755-1760	1.10 [0.91;1.32]	1.08 [0.64;1.80]
birth cohort 1760-1765	1.16 [0.98;1.38]	0.95 [0.60;1.51]
birth cohort 1765-1770	1.11 [0.93;1.31]	0.71 [0.44;1.13]
birth cohort 1770-1775	1.06 [0.89;1.26]	0.90 [0.56;1.46]
birth cohort 1775-1780	1.05 [0.88;1.23]	1.06 [0.67;1.68]
birth cohort 1780-1785	1.16 [0.97;1.37]	0.99 [0.62;1.58]
birth cohort 1785-1790	1.13 [0.96;1.33]	1.14 [0.74;1.73]
birth cohort 1790-1795	1.03 [0.88;1.21]	1.36 [0.89;2.06]
birth cohort 1795-1800	1.03 [0.87;1.19]	1.12 [0.74;1.67]
birth cohort 1800-1805	0.98 [0.83;1.13]	1.03 [0.68;1.54]
birth cohort 1805-1810	0.99 [0.84;1.15]	0.99 [0.66;1.47]
birth cohort 1810-1815	1.02 [0.87;1.18]	1.07 [0.71;1.57]
birth cohort 1815-1820	1.07 [0.92;1.24]	0.90 [0.60;1.34]
birth cohort 1820-1825	1.09 [0.93;1.27]	0.81 [0.55;1.19]
birth cohort 1825-1830	1.05 [0.90;1.21]	0.81 [0.55;1.21]
birth cohort 1830-1835	1.07 [0.92;1.24]	0.83 [0.56;1.23]
birth cohort 1835-1840	1.07 [0.91;1.24]	0.81 [0.54;1.20]
birth cohort 1840-1845	1.04 [0.90;1.21]	0.82 [0.55;1.22]
birth cohort 1845-1850	1.05 [0.90;1.22]	0.85 [0.57;1.27]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
Intercept	3.76 [3.21;4.44]	0.94 [0.62;1.44]
last born	0.98 [0.96;1.00]	1.01 [0.96;1.06]
male	1.04 [1.03;1.05]	1.05 [1.01;1.09]
maternal loss 0-1	1.06 [0.96;1.18]	6.41 [5.06;8.11]
maternal loss 1-5	1.00 [0.94;1.05]	2.80 [2.46;3.19]
maternal loss 10-15	0.96 [0.92;1.00]	2.24 [2.02;2.49]
maternal loss 15-20	0.96 [0.92;1.00]	1.96 [1.78;2.15]
maternal loss 20-25	0.93 [0.90;0.96]	1.58 [1.44;1.72]
maternal loss 25-30	0.97 [0.94;1.00]	1.37 [1.27;1.48]
maternal loss 30-35	0.98 [0.95;1.00]	1.28 [1.19;1.37]
maternal loss 35-40	1.00 [0.97;1.02]	1.16 [1.08;1.24]
maternal loss 40-45	0.98 [0.96;1.00]	1.08 [1.01;1.15]
maternal loss 5-10	0.97 [0.93;1.02]	2.38 [2.14;2.66]
maternalage factor 10-20	1.04 [0.99;1.11]	1.05 [0.89;1.26]
maternalage factor 35-59	1.05 [1.03;1.07]	1.08 [1.03;1.14]
nr siblings	1.03 [1.02;1.03]	1.04 [1.03;1.06]
older siblings 1	1.01 [0.99;1.02]	0.97 [0.92;1.02]
older siblings 2	1.01 [0.99;1.03]	0.99 [0.92;1.05]
older siblings 3	1.00 [0.98;1.03]	0.98 [0.91;1.06]
older siblings 4	0.97 [0.94;1.00]	0.96 [0.88;1.05]
older siblings 5+	0.97 [0.94;1.00]	0.91 [0.83;1.00]
paternal loss 0-1	1.05 [0.98;1.13]	2.39 [1.99;2.87]
paternal loss 1-5	1.02 [0.97;1.07]	1.97 [1.74;2.22]
paternal loss 10-15	0.97 [0.94;1.01]	1.69 [1.55;1.86]
paternal loss 15-20	0.92 [0.89;0.96]	1.53 [1.40;1.67]
paternal loss 20-25	0.97 [0.94;1.00]	1.38 [1.27;1.50]
paternal loss 25-30	0.97 [0.95;1.00]	1.26 [1.16;1.36]
paternal loss 30-35	0.98 [0.95;1.01]	1.17 [1.09;1.27]
paternal loss 35-40	1.02 [0.99;1.05]	1.13 [1.05;1.21]
paternal loss 40-45	1.03 [1.00;1.05]	1.04 [0.96;1.12]
paternal loss 5-10	0.97 [0.93;1.01]	2.00 [1.80;2.22]
paternalage mean	1.01 [0.99;1.02]	0.97 [0.94;1.01]

4.2.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

This model did not contain a within family paternal age predictor.

4.3 *m3*: Sibling comparison, linear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a linear effect for paternal age differences among siblings.

4.3.1 Model summary

Data: 56663 individuals nested in 14746 mother-father dyads.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

4.3.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

4.3.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.82 [0.78;0.85]	0.36 [0.34;0.37]

4.3.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1750-1755	0.84 [0.67;1.06]	1.45 [0.83;2.54]
birth cohort 1755-1760	1.10 [0.91;1.32]	1.09 [0.68;1.80]
birth cohort 1760-1765	1.16 [0.98;1.37]	0.96 [0.63;1.49]
birth cohort 1765-1770	1.10 [0.93;1.31]	0.72 [0.47;1.10]
birth cohort 1770-1775	1.06 [0.89;1.26]	0.91 [0.58;1.40]
birth cohort 1775-1780	1.05 [0.88;1.24]	1.07 [0.69;1.61]
birth cohort 1780-1785	1.16 [0.99;1.37]	1.00 [0.66;1.53]
birth cohort 1785-1790	1.13 [0.96;1.34]	1.15 [0.77;1.68]
birth cohort 1790-1795	1.03 [0.88;1.20]	1.39 [0.94;1.96]
birth cohort 1795-1800	1.02 [0.88;1.20]	1.14 [0.78;1.65]
birth cohort 1800-1805	0.97 [0.84;1.14]	1.05 [0.71;1.51]
birth cohort 1805-1810	0.99 [0.85;1.15]	1.00 [0.69;1.44]
birth cohort 1810-1815	1.01 [0.87;1.18]	1.08 [0.74;1.54]
birth cohort 1815-1820	1.07 [0.92;1.24]	0.91 [0.63;1.30]
birth cohort 1820-1825	1.09 [0.94;1.27]	0.82 [0.56;1.17]
birth cohort 1825-1830	1.04 [0.90;1.21]	0.82 [0.58;1.18]
birth cohort 1830-1835	1.07 [0.92;1.25]	0.84 [0.58;1.19]
birth cohort 1835-1840	1.07 [0.92;1.24]	0.82 [0.57;1.16]
birth cohort 1840-1845	1.05 [0.90;1.21]	0.83 [0.58;1.18]
birth cohort 1845-1850	1.05 [0.91;1.23]	0.86 [0.60;1.22]
Intercept	3.74 [3.19;4.39]	0.93 [0.65;1.38]
last born	0.98 [0.96;1.00]	1.01 [0.96;1.06]
male	1.04 [1.03;1.05]	1.05 [1.01;1.09]
maternal loss 0-1	1.07 [0.96;1.18]	6.40 [5.08;8.15]
maternal loss 1-5	1.00 [0.95;1.06]	2.78 [2.42;3.20]
maternal loss 10-15	0.96 [0.93;1.01]	2.23 [1.99;2.50]
maternal loss 15-20	0.96 [0.93;1.00]	1.96 [1.78;2.16]
maternal loss 20-25	0.93 [0.90;0.97]	1.58 [1.45;1.73]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
maternal loss 25-30	0.97 [0.94;1.00]	1.37 [1.26;1.49]
maternal loss 30-35	0.98 [0.95;1.01]	1.28 [1.19;1.38]
maternal loss 35-40	1.00 [0.98;1.03]	1.15 [1.08;1.23]
maternal loss 40-45	0.98 [0.96;1.00]	1.08 [1.01;1.16]
maternal loss 5-10	0.98 [0.94;1.03]	2.37 [2.11;2.67]
maternalage factor 10-20	1.04 [0.98;1.10]	1.06 [0.89;1.27]
maternalage factor 35-59	1.07 [1.04;1.09]	1.07 [1.01;1.14]
nr siblings	1.02 [1.02;1.03]	1.05 [1.03;1.06]
older siblings 1	1.02 [0.99;1.04]	0.96 [0.90;1.02]
older siblings 2	1.03 [1.00;1.06]	0.96 [0.89;1.05]
older siblings 3	1.04 [1.00;1.07]	0.95 [0.86;1.05]
older siblings 4	1.01 [0.97;1.06]	0.92 [0.81;1.04]
older siblings 5+	1.03 [0.97;1.09]	0.86 [0.73;1.01]
paternal loss 0-1	1.06 [0.99;1.14]	2.38 [1.98;2.85]
paternal loss 1-5	1.03 [0.98;1.08]	1.95 [1.73;2.21]
paternal loss 10-15	0.98 [0.94;1.02]	1.69 [1.53;1.86]
paternal loss 15-20	0.93 [0.90;0.96]	1.53 [1.39;1.67]
paternal loss 20-25	0.97 [0.94;1.00]	1.38 [1.27;1.50]
paternal loss 25-30	0.98 [0.95;1.01]	1.25 [1.15;1.35]
paternal loss 30-35	0.99 [0.96;1.01]	1.17 [1.09;1.26]
paternal loss 35-40	1.02 [1.00;1.05]	1.13 [1.05;1.21]
paternal loss 40-45	1.03 [1.00;1.05]	1.04 [0.97;1.12]
paternal loss 5-10	0.98 [0.94;1.03]	1.99 [1.79;2.21]
paternalage	0.95 [0.91;0.98]	1.05 [0.95;1.17]
paternalage mean	1.06 [1.02;1.10]	0.93 [0.83;1.03]

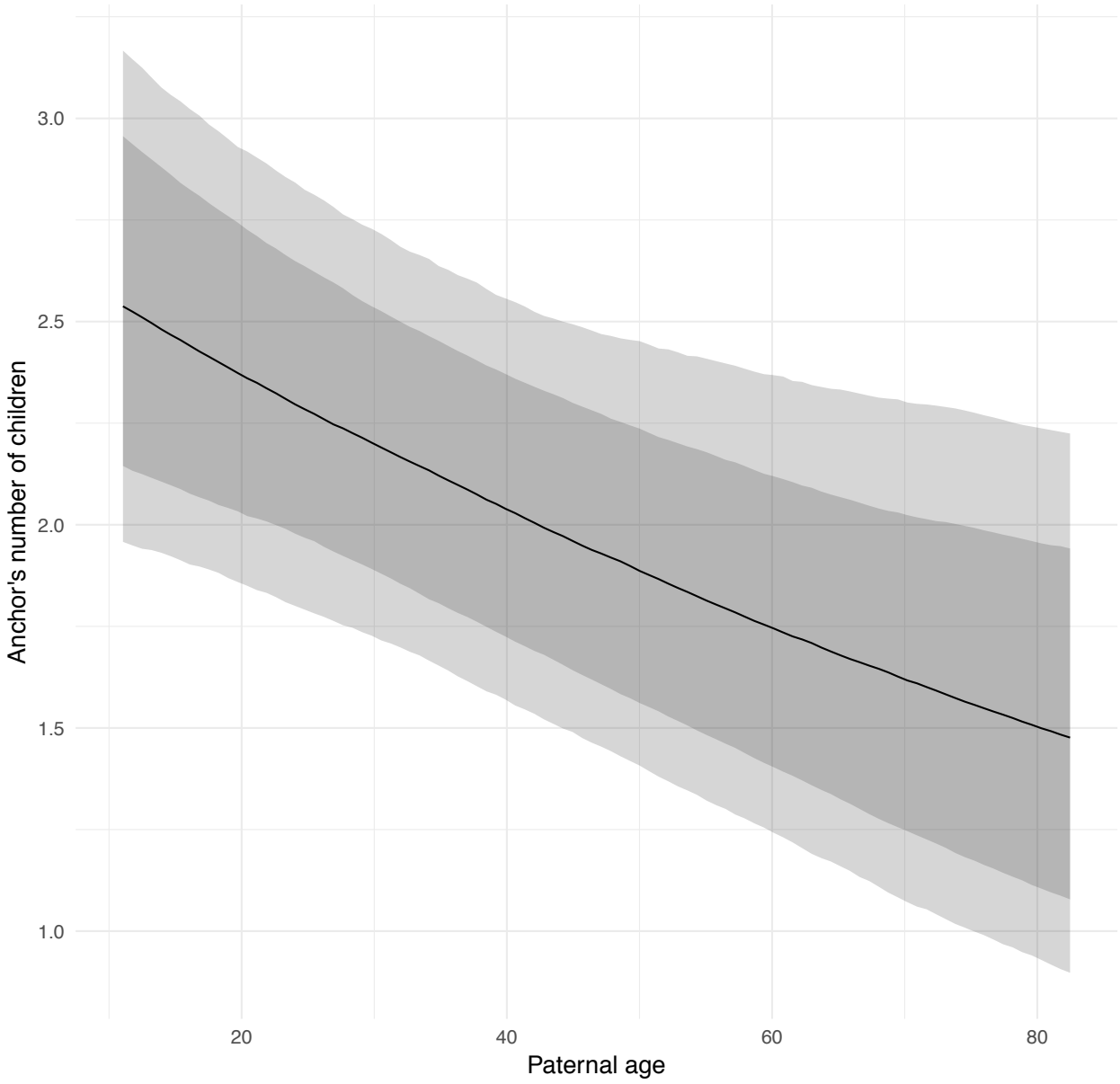
4.3.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-7.29	[-13.40; -1.07]	[-11.15; -3.33]

4.3.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



4.4 *m4*: Sibling comparison, nonlinear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a possibly nonlinear effect for paternal age differences among siblings.

4.4.1 Model summary

Data: 56663 individuals nested in 14746 mother-father dyads.

Formula (Wilkinson notation): children ~ s(paternalage) + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: hurdle_poisson
- **link**: log

4.4.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd
student_t(3, 0, 10)	sds

4.4.3 Group-level effects

Component	Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.82 [0.79;0.85]	0.36 [0.35;0.37]

4.4.3.1 Splines

Effect	Hurdle Estimate	Zero-truncated Poisson Estimate
sds(spaternalage_1)	0.36 [0.01;1.30]	0.28 [0.03;0.75]

4.4.4 Population-level effects

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1750-1755	0.84 [0.68;1.06]	1.44 [0.86;2.50]
birth cohort 1755-1760	1.10 [0.91;1.33]	1.08 [0.68;1.73]
birth cohort 1760-1765	1.16 [0.98;1.39]	0.96 [0.63;1.46]
birth cohort 1765-1770	1.10 [0.93;1.33]	0.71 [0.46;1.09]
birth cohort 1770-1775	1.06 [0.90;1.28]	0.90 [0.58;1.39]
birth cohort 1775-1780	1.05 [0.89;1.25]	1.06 [0.70;1.56]
birth cohort 1780-1785	1.17 [0.99;1.38]	0.99 [0.66;1.50]
birth cohort 1785-1790	1.14 [0.97;1.34]	1.15 [0.78;1.68]
birth cohort 1790-1795	1.03 [0.88;1.21]	1.38 [0.95;1.99]
birth cohort 1795-1800	1.02 [0.87;1.20]	1.13 [0.79;1.59]
birth cohort 1800-1805	0.98 [0.84;1.15]	1.04 [0.73;1.49]

Effect	Hurdle Odds ratio	Zero-truncated Poisson Hazard ratio
birth cohort 1805-1810	0.99 [0.85;1.16]	1.00 [0.70;1.40]
birth cohort 1810-1815	1.02 [0.88;1.19]	1.07 [0.74;1.50]
birth cohort 1815-1820	1.07 [0.92;1.25]	0.91 [0.64;1.26]
birth cohort 1820-1825	1.09 [0.94;1.27]	0.82 [0.57;1.14]
birth cohort 1825-1830	1.05 [0.90;1.22]	0.82 [0.58;1.14]
birth cohort 1830-1835	1.07 [0.92;1.25]	0.83 [0.58;1.16]
birth cohort 1835-1840	1.07 [0.92;1.25]	0.81 [0.58;1.13]
birth cohort 1840-1845	1.05 [0.90;1.23]	0.83 [0.58;1.15]
birth cohort 1845-1850	1.06 [0.91;1.24]	0.86 [0.60;1.20]
Intercept	3.15 [2.54;3.85]	1.09 [0.64;1.90]
last born	0.98 [0.96;1.00]	1.01 [0.96;1.06]
male	1.04 [1.03;1.06]	1.05 [1.01;1.09]
maternal loss 0-1	1.07 [0.96;1.19]	6.37 [5.03;8.00]
maternal loss 1-5	1.00 [0.95;1.06]	2.78 [2.43;3.17]
maternal loss 10-15	0.97 [0.93;1.01]	2.24 [2.00;2.49]
maternal loss 15-20	0.97 [0.93;1.00]	1.96 [1.78;2.16]
maternal loss 20-25	0.93 [0.90;0.97]	1.58 [1.44;1.73]
maternal loss 25-30	0.97 [0.95;1.00]	1.37 [1.27;1.49]
maternal loss 30-35	0.98 [0.95;1.01]	1.28 [1.18;1.38]
maternal loss 35-40	1.00 [0.98;1.03]	1.15 [1.08;1.24]
maternal loss 40-45	0.98 [0.96;1.00]	1.08 [1.01;1.15]
maternal loss 5-10	0.98 [0.94;1.03]	2.38 [2.12;2.66]
maternalage factor 10-20	1.04 [0.98;1.10]	1.06 [0.89;1.27]
maternalage factor 35-59	1.06 [1.04;1.09]	1.07 [1.01;1.14]
nr siblings	1.02 [1.02;1.03]	1.05 [1.03;1.06]
older siblings 1	1.01 [0.99;1.04]	0.96 [0.91;1.03]
older siblings 2	1.02 [1.00;1.05]	0.97 [0.89;1.06]
older siblings 3	1.03 [0.99;1.07]	0.96 [0.86;1.07]
older siblings 4	1.00 [0.96;1.05]	0.93 [0.82;1.07]
older siblings 5+	1.02 [0.96;1.08]	0.88 [0.73;1.03]
paternal loss 0-1	1.06 [0.99;1.14]	2.38 [1.98;2.85]
paternal loss 1-5	1.03 [0.98;1.08]	1.95 [1.73;2.20]
paternal loss 10-15	0.98 [0.94;1.02]	1.68 [1.53;1.85]
paternal loss 15-20	0.93 [0.90;0.96]	1.53 [1.39;1.66]
paternal loss 20-25	0.97 [0.94;1.00]	1.38 [1.27;1.50]
paternal loss 25-30	0.98 [0.95;1.01]	1.26 [1.16;1.36]
paternal loss 30-35	0.98 [0.95;1.01]	1.17 [1.09;1.27]
paternal loss 35-40	1.02 [0.99;1.04]	1.13 [1.05;1.21]
paternal loss 40-45	1.03 [1.00;1.05]	1.04 [0.97;1.13]
paternal loss 5-10	0.98 [0.94;1.02]	1.99 [1.81;2.21]
paternalage mean	1.06 [1.01;1.10]	0.93 [0.83;1.04]
spaternalage	0.95 [0.87;1.03]	1.03 [0.88;1.23]

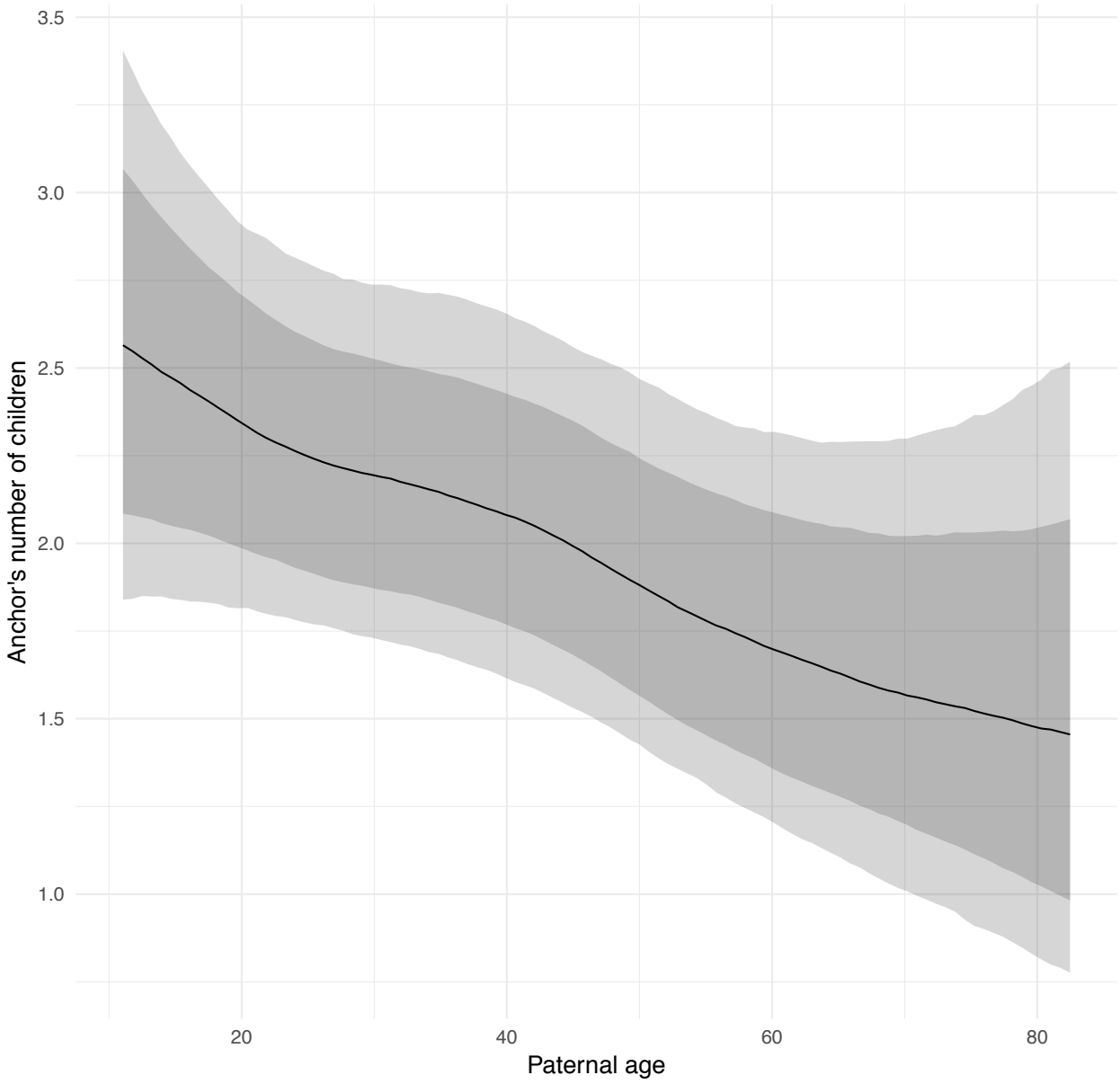
4.4.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-4.43	[-11.65; 2.93]	[-9.10; 0.47]

4.4.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



4.5 Main model comparison

We compare the four models using an approximate leave-one-out cross-validation information criterion as implemented in `brms` and `loo` and the Watanabe-Akaike information criterion.

4.5.1 Approximate leave-one-out (LOO) cross-validation

	LOOIC	SE
m1	191798	758.7
m2	185251	723
m3	185226	722.9
m4	185248	723
m1 - m2	6547	217.8
m1 - m3	6572	218.1
m1 - m4	6550	218.3
m2 - m3	25.39	15.36
m2 - m4	3.12	16.38
m3 - m4	-22.27	14.58

4.5.2 Watanabe-Akaike information criterion

	WAIC	SE
m1	191798	758.7
m2	184258	715.3
m3	184250	715.4
m4	184259	715.4
m1 - m2	7540	219.5
m1 - m3	7548	219.8
m1 - m4	7539	219.9
m2 - m3	7.65	11.75
m2 - m4	-1.11	12.89
m3 - m4	-8.76	10.85

4.6 *e1*: Selective episode: offspring survival of the first year

In the first selective episode model, we tested how much of the paternal age effect happens in the first selective episode, i.e. in the offspring's survival of the first year.

4.6.1 Model summary

Data: 56010 individuals nested in 14708 mother-father dyads.

Formula (Wilkinson notation): survive1y ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: bernoulli
- **link**: cauchit

4.6.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

4.6.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	1.41 [1.33;1.49]

4.6.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1750-1755	0.83 [0.11; 12.83]
birth cohort 1755-1760	0.35 [0.08; 1.60]
birth cohort 1760-1765	0.25 [0.06; 0.87]
birth cohort 1765-1770	0.77 [0.17; 3.51]
birth cohort 1770-1775	0.67 [0.15; 3.03]
birth cohort 1775-1780	0.14 [0.04; 0.44]
birth cohort 1780-1785	0.14 [0.04; 0.42]
birth cohort 1785-1790	0.19 [0.05; 0.59]
birth cohort 1790-1795	0.16 [0.04; 0.48]
birth cohort 1795-1800	0.24 [0.06; 0.71]
birth cohort 1800-1805	0.23 [0.06; 0.69]
birth cohort 1805-1810	0.19 [0.05; 0.57]
birth cohort 1810-1815	0.17 [0.05; 0.52]
birth cohort 1815-1820	0.17 [0.05; 0.52]
birth cohort 1820-1825	0.27 [0.08; 0.79]
birth cohort 1825-1830	0.31 [0.08; 0.91]
birth cohort 1830-1835	0.26 [0.07; 0.80]
birth cohort 1835-1840	0.29 [0.08; 0.84]
birth cohort 1840-1845	0.40 [0.11; 1.21]

Effect	Hurdle Odds ratio
birth cohort 1845-1850	0.45 [0.13; 1.33]
Intercept	165.45 [54.03;604.23]
last born	0.93 [0.83; 1.04]
male	0.69 [0.63; 0.75]
maternal loss 0-1	0.03 [0.02; 0.04]
maternal loss 1-5	0.30 [0.24; 0.38]
maternal loss 10-15	0.58 [0.47; 0.71]
maternal loss 15-20	0.64 [0.52; 0.79]
maternal loss 20-25	0.65 [0.53; 0.78]
maternal loss 25-30	0.64 [0.53; 0.77]
maternal loss 30-35	0.73 [0.62; 0.87]
maternal loss 35-40	0.84 [0.72; 1.00]
maternal loss 40-45	0.93 [0.77; 1.10]
maternal loss 5-10	0.39 [0.32; 0.48]
maternalage factor 10-20	0.91 [0.63; 1.35]
maternalage factor 35-59	0.81 [0.71; 0.92]
nr siblings	0.78 [0.75; 0.80]
older siblings 1	1.83 [1.59; 2.10]
older siblings 2	2.50 [2.09; 2.96]
older siblings 3	2.86 [2.35; 3.52]
older siblings 4	4.02 [3.15; 5.19]
older siblings 5+	7.31 [5.28; 10.31]
paternal loss 0-1	0.37 [0.27; 0.51]
paternal loss 1-5	0.63 [0.50; 0.81]
paternal loss 10-15	0.74 [0.60; 0.91]
paternal loss 15-20	0.82 [0.67; 1.00]
paternal loss 20-25	0.92 [0.76; 1.12]
paternal loss 25-30	0.93 [0.77; 1.13]
paternal loss 30-35	0.87 [0.73; 1.05]
paternal loss 35-40	1.00 [0.84; 1.20]
paternal loss 40-45	1.10 [0.90; 1.34]
paternal loss 5-10	0.67 [0.54; 0.84]
paternalage	0.29 [0.24; 0.36]
paternalage mean	3.53 [2.82; 4.45]

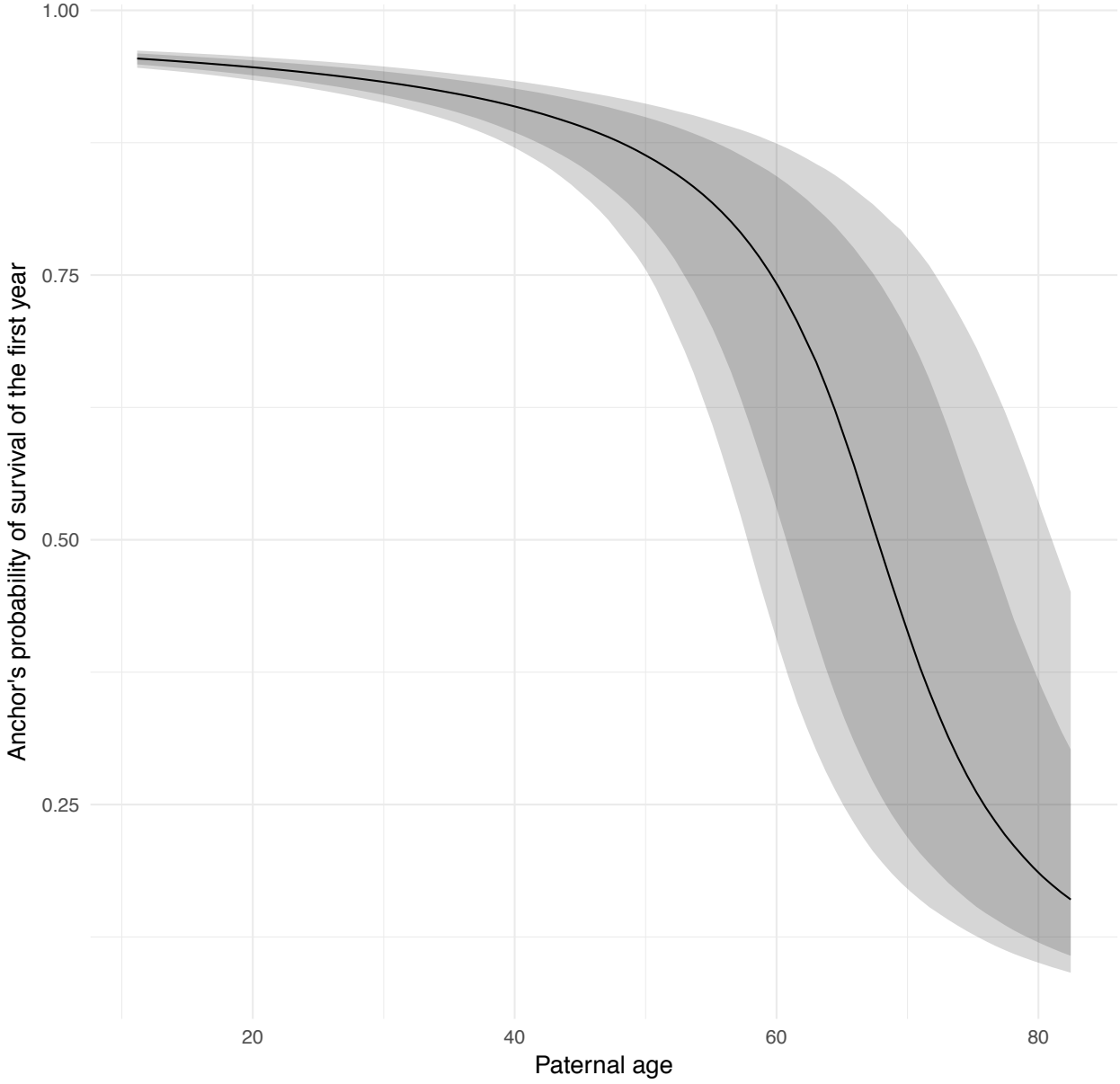
4.6.5 Paternal age effect

This is the effect of 10 years of paternal age within families on probability of survival of the first year, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-1.82	[-3.14;-1.08]	[-2.63;-1.28]

4.6.5.1 Marginal effect plot

Paternal age effect on probability of survival of the first year The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



5 20th-century Sweden

5.1 *m1*: No sibling comparison

Here, we ignore the pedigree structure of the data to see whether it matters for the estimation of the paternal age effect.

5.1.1 Model summary

Data: 1408177 individuals.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born.

- **family**: poisson
- **link**: log

5.1.2 Priors

prior	class
normal(0,5)	b

5.1.3 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1950-1955	1.00 [1.00;1.00]
birth cohort 1955-1960	1.00 [1.00;1.01]
Intercept	2.08 [2.06;2.10]
last born	1.01 [1.01;1.01]
male	0.94 [0.94;0.94]
maternal loss 0-1	0.82 [0.71;0.94]
maternal loss 1-5	0.95 [0.89;1.00]
maternal loss 10-15	0.99 [0.98;1.01]
maternal loss 15-20	1.00 [0.98;1.01]
maternal loss 20-25	1.01 [1.00;1.02]
maternal loss 25-30	1.01 [1.00;1.01]
maternal loss 30-35	1.01 [1.00;1.01]
maternal loss 35-40	1.00 [0.99;1.00]
maternal loss 40-45	1.00 [0.99;1.01]
maternal loss 5-10	0.99 [0.96;1.02]
maternal loss unclear	0.98 [0.98;0.98]
maternalage factor 14-20	1.06 [1.05;1.06]
maternalage factor 3-56	1.00 [0.99;1.00]
nr siblings	1.04 [1.04;1.04]
older siblings 1	1.02 [1.01;1.02]
older siblings 2	1.02 [1.01;1.02]
older siblings 3	1.01 [1.00;1.02]
older siblings 4	0.99 [0.98;1.00]
older siblings 5+	0.94 [0.93;0.95]
paternal loss 0-1	1.12 [1.00;1.24]

Effect	Hurdle Odds ratio
paternal loss 1-5	1.03 [1.00;1.07]
paternal loss 10-15	1.00 [0.99;1.01]
paternal loss 15-20	1.00 [1.00;1.01]
paternal loss 20-25	1.00 [0.99;1.00]
paternal loss 25-30	1.00 [1.00;1.01]
paternal loss 30-35	1.00 [1.00;1.00]
paternal loss 35-40	1.00 [0.99;1.00]
paternal loss 40-45	0.99 [0.99;1.00]
paternal loss 5-10	0.98 [0.97;1.00]
paternal loss unclear	0.95 [0.94;0.95]
paternalage	0.95 [0.95;0.96]

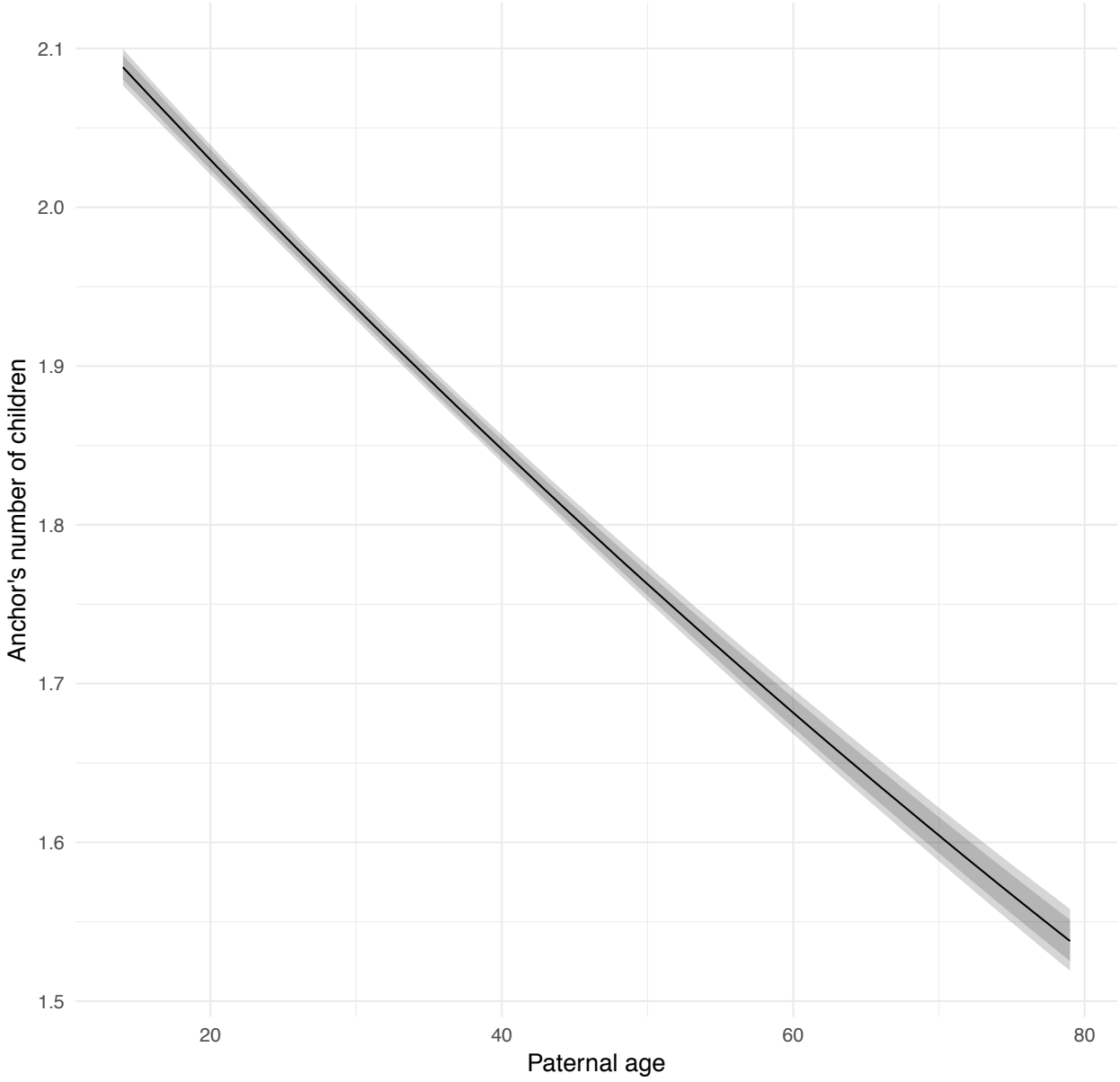
5.1.4 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-4.60	[-4.83;-4.36]	[-4.75;-4.44]

5.1.4.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



5.2 *m2*: Sibling comparison, no paternal age effect

Here, we compared siblings by including a random intercept for the family, but we modelled no effect for paternal age differences among siblings.

5.2.1 Model summary

Data: 1408177 individuals nested in 884975 mother-father dyads.

Formula (Wilkinson notation): children ~ birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: poisson
- **link**: log

5.2.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

5.2.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.00 [0.00;0.01]

5.2.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1950-1955	1.00 [1.00;1.00]
birth cohort 1955-1960	1.00 [1.00;1.00]
Intercept	2.07 [2.05;2.09]
last born	1.01 [1.01;1.02]
male	0.94 [0.94;0.94]
maternal loss 0-1	0.82 [0.71;0.94]
maternal loss 1-5	0.94 [0.89;1.00]
maternal loss 10-15	0.99 [0.98;1.01]
maternal loss 15-20	1.00 [0.98;1.01]
maternal loss 20-25	1.01 [1.00;1.02]
maternal loss 25-30	1.01 [1.00;1.01]
maternal loss 30-35	1.01 [1.00;1.01]
maternal loss 35-40	1.00 [0.99;1.00]
maternal loss 40-45	1.00 [0.99;1.00]
maternal loss 5-10	0.99 [0.96;1.01]
maternal loss unclear	0.98 [0.98;0.98]
maternalage factor 14-20	1.06 [1.06;1.07]
maternalage factor 3-56	0.99 [0.99;0.99]
nr siblings	1.05 [1.04;1.05]
older siblings 1	1.00 [1.00;1.00]

Effect	Hurdle Odds ratio
older siblings 2	0.99 [0.98;0.99]
older siblings 3	0.97 [0.96;0.97]
older siblings 4	0.94 [0.93;0.95]
older siblings 5+	0.87 [0.86;0.88]
paternal loss 0-1	1.12 [1.02;1.24]
paternal loss 1-5	1.03 [0.99;1.07]
paternal loss 10-15	0.99 [0.98;1.00]
paternal loss 15-20	1.00 [0.99;1.01]
paternal loss 20-25	1.00 [0.99;1.00]
paternal loss 25-30	1.00 [1.00;1.01]
paternal loss 30-35	1.00 [0.99;1.00]
paternal loss 35-40	0.99 [0.99;1.00]
paternal loss 40-45	0.99 [0.99;1.00]
paternal loss 5-10	0.98 [0.96;1.00]
paternal loss unclear	0.95 [0.94;0.95]
paternalage mean	0.96 [0.95;0.96]

5.2.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

This model did not contain a within family paternal age predictor.

5.3 *m3*: Sibling comparison, linear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a linear effect for paternal age differences among siblings.

5.3.1 Model summary

Data: 1408177 individuals nested in 884975 mother-father dyads.

Formula (Wilkinson notation): children ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: poisson
- **link**: log

5.3.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

5.3.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.00 [0.00;0.01]

5.3.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1950-1955	1.00 [1.00;1.00]
birth cohort 1955-1960	1.00 [1.00;1.01]
Intercept	2.08 [2.06;2.10]
last born	1.01 [1.01;1.01]
male	0.94 [0.94;0.94]
maternal loss 0-1	0.82 [0.71;0.95]
maternal loss 1-5	0.95 [0.89;1.00]
maternal loss 10-15	0.99 [0.98;1.01]
maternal loss 15-20	1.00 [0.98;1.01]
maternal loss 20-25	1.01 [1.00;1.02]
maternal loss 25-30	1.01 [1.00;1.01]
maternal loss 30-35	1.01 [1.00;1.01]
maternal loss 35-40	1.00 [0.99;1.00]
maternal loss 40-45	1.00 [0.99;1.00]
maternal loss 5-10	0.99 [0.96;1.02]
maternal loss unclear	0.98 [0.98;0.98]
maternalage factor 14-20	1.06 [1.05;1.06]
maternalage factor 3-56	1.00 [1.00;1.00]
nr siblings	1.04 [1.04;1.04]
older siblings 1	1.02 [1.01;1.02]
older siblings 2	1.02 [1.02;1.03]
older siblings 3	1.02 [1.01;1.03]
older siblings 4	1.00 [0.99;1.02]
older siblings 5+	0.95 [0.93;0.97]
paternal loss 0-1	1.12 [1.00;1.25]
paternal loss 1-5	1.03 [1.00;1.07]
paternal loss 10-15	1.00 [0.99;1.01]
paternal loss 15-20	1.00 [1.00;1.01]
paternal loss 20-25	1.00 [0.99;1.00]
paternal loss 25-30	1.00 [1.00;1.01]
paternal loss 30-35	1.00 [1.00;1.01]
paternal loss 35-40	1.00 [0.99;1.00]
paternal loss 40-45	0.99 [0.99;1.00]
paternal loss 5-10	0.98 [0.97;1.00]
paternal loss unclear	0.95 [0.94;0.95]
paternalage	0.95 [0.94;0.96]
paternalage mean	1.01 [1.00;1.01]

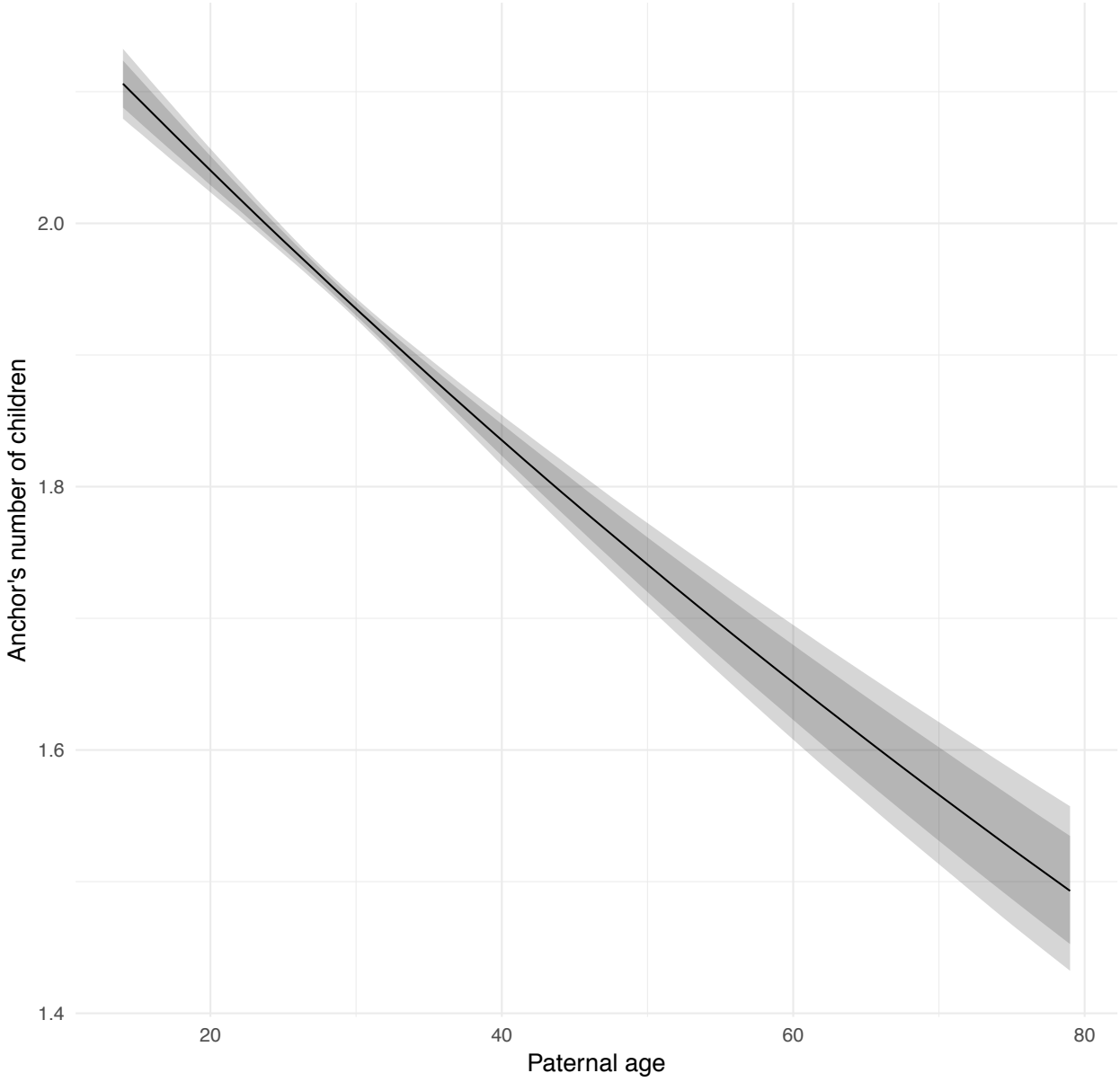
5.3.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-5.15	[-5.94;-4.36]	[-5.67;-4.63]

5.3.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



5.4 *m4*: Sibling comparison, nonlinear paternal age effect

Here, we compared siblings by including a random intercept for the family, and we modelled a possibly nonlinear effect for paternal age differences among siblings.

5.4.1 Model summary

Data: 1408177 individuals nested in 884975 mother-father dyads.

Formula (Wilkinson notation): children ~ s(paternalage) + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: poisson
- **link**: log

5.4.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd
student_t(3, 0, 10)	sds

5.4.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	0.00 [0.00;0.01]

5.4.3.1 Splines

Effect	Zero-truncated Poisson Estimate
sds(spaternalage_1)	0.13 [0.06;0.37]

5.4.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1950-1955	1.00 [1.00;1.00]
birth cohort 1955-1960	1.00 [1.00;1.01]
Intercept	1.75 [1.71;1.80]
last born	1.01 [1.01;1.01]
male	0.94 [0.94;0.94]
maternal loss 0-1	0.82 [0.71;0.94]
maternal loss 1-5	0.94 [0.89;1.00]
maternal loss 10-15	0.99 [0.98;1.01]
maternal loss 15-20	0.99 [0.98;1.01]
maternal loss 20-25	1.01 [1.00;1.02]
maternal loss 25-30	1.00 [1.00;1.01]

Effect	Hurdle Odds ratio
maternal loss 30-35	1.00 [1.00;1.01]
maternal loss 35-40	1.00 [0.99;1.00]
maternal loss 40-45	1.00 [0.99;1.00]
maternal loss 5-10	0.99 [0.96;1.02]
maternal loss unclear	0.98 [0.98;0.98]
maternalage factor 14-20	1.05 [1.04;1.05]
maternalage factor 3-56	0.99 [0.99;1.00]
nr siblings	1.04 [1.04;1.04]
older siblings 1	1.02 [1.02;1.03]
older siblings 2	1.03 [1.02;1.04]
older siblings 3	1.02 [1.01;1.03]
older siblings 4	1.01 [0.99;1.02]
older siblings 5+	0.95 [0.94;0.97]
paternal loss 0-1	1.11 [1.00;1.22]
paternal loss 1-5	1.03 [0.99;1.06]
paternal loss 10-15	0.99 [0.98;1.00]
paternal loss 15-20	1.00 [0.99;1.01]
paternal loss 20-25	0.99 [0.99;1.00]
paternal loss 25-30	1.00 [0.99;1.01]
paternal loss 30-35	1.00 [0.99;1.00]
paternal loss 35-40	1.00 [0.99;1.00]
paternal loss 40-45	0.99 [0.99;1.00]
paternal loss 5-10	0.98 [0.96;0.99]
paternal loss unclear	0.94 [0.94;0.95]
paternalage mean	1.01 [1.00;1.02]
spaternalage	0.96 [0.92;1.00]

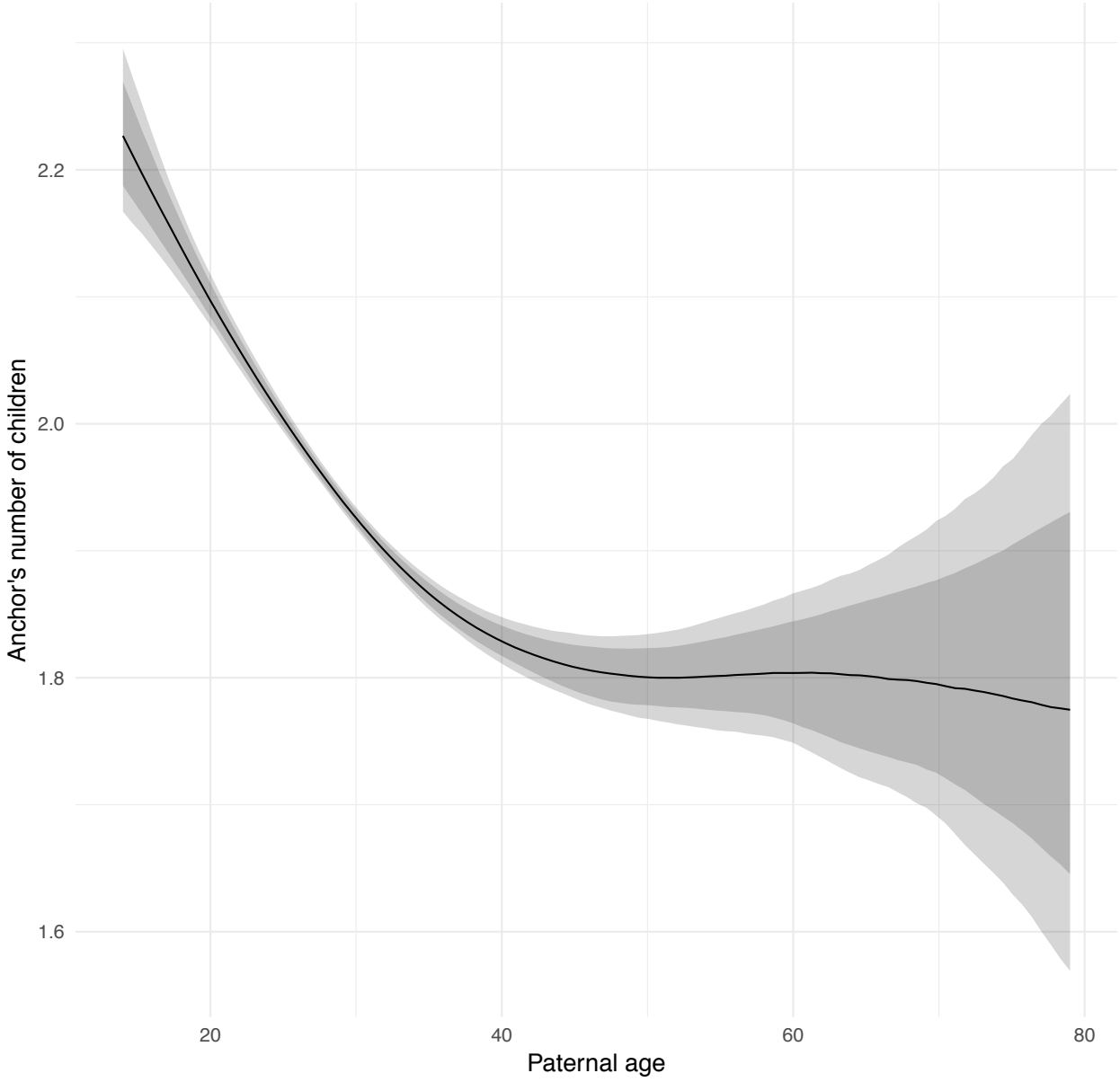
5.4.5 Paternal age effect

This is the effect of 10 years of paternal age within families on number of children, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-6.88	[-7.62;-6.03]	[-7.40;-6.32]

5.4.5.1 Marginal effect plot

Paternal age effect on number of children The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



5.5 Main model comparison

Because of computational limitations, we could not compute LOO and WAIC for the complete dataset models in this population. Instead, we computed the main models and their LOOs and WAICs using a randomly drawn subset of 100,000 families (the same that was used for the robustness analyses).

We compare the four models using an approximate leave-one-out cross-validation information criterion as implemented in `brms` and `loo` and the Watanabe-Akaike information criterion.

5.5.1 Approximate leave-one-out (LOO) cross-validation

	LOOIC	SE
m1	415187	388.2
m2	415196	388
m3	415186	388.1
m4	415165	387.9
m1 - m2	-9.13	6.31
m1 - m3	0.93	0.42
m1 - m4	22.09	9.37
m2 - m3	10.07	6.47
m2 - m4	31.22	11.3
m3 - m4	21.15	9.36

5.5.2 Watanabe-Akaike information criterion

	WAIC	SE
m1	415187	388.2
m2	415196	388
m3	415186	388.1
m4	415165	387.9
m1 - m2	-9.1	6.31
m1 - m3	0.89	0.42
m1 - m4	22.04	9.37
m2 - m3	9.98	6.47
m2 - m4	31.14	11.3
m3 - m4	21.16	9.36

5.6 *e1*: Selective episode: offspring survival of the first year

In the first selective episode model, we tested how much of the paternal age effect happens in the first selective episode, i.e. in the offspring's survival of the first year.

5.6.1 Model summary

Data: 363744 individuals nested in 200000 mother-father dyads.

Formula (Wilkinson notation): survive1y ~ paternalage + birth_cohort + male + maternalage.factor + paternalage.mean + paternal_loss + maternal_loss + older_siblings + nr.siblings + last_born + (1 | idParents).

- **family**: bernoulli
- **link**: cauchit

5.6.2 Priors

prior	class
normal(0,5)	b
student_t(3, 0, 5)	sd

5.6.3 Group-level effects

Component	Effect	Zero-truncated Poisson Estimate
idParents	sd(Intercept)	21.41 [19.76;22.96]

5.6.4 Population-level effects

Effect	Hurdle Odds ratio
birth cohort 1970-1975	0.0e+00 [0.0e+00;1.7e-01]
birth cohort 1975-1980	4.0e-02 [0.0e+00;2.1e+00]
birth cohort 1980-1985	1.1e+02 [1.4e+00;7.2e+03]
birth cohort 1985-1990	5.0e+01 [6.6e-01;3.7e+03]
birth cohort 1990-1995	3.9e+02 [5.6e+00;3.3e+04]
birth cohort 1995-2000	2.6e+06 [7.5e+03;1.0e+09]
Intercept	5.2e+41 [7.2e+37;7.2e+45]
last born	0.0e+00 [0.0e+00;0.0e+00]
male	1.0e-02 [0.0e+00;6.0e-02]
maternal loss 0-1	1.0e-02 [0.0e+00;2.6e+02]
maternal loss 1-5	6.7e+00 [0.0e+00;7.3e+04]
maternal loss 10-15	7.0e-01 [0.0e+00;3.6e+03]
maternal loss 15-20	0.0e+00 [0.0e+00;1.7e+00]
maternal loss 20-25	0.0e+00 [0.0e+00;1.4e+00]
maternal loss 25-30	1.0e-02 [0.0e+00;6.9e+00]
maternal loss 30-35	5.0e-02 [0.0e+00;5.4e+01]
maternal loss 35-40	2.0e-01 [0.0e+00;7.1e+02]
maternal loss 40-45	2.6e+00 [0.0e+00;2.5e+04]
maternal loss 5-10	4.7e+00 [0.0e+00;3.8e+04]

Effect	Hurdle Odds ratio
maternalage factor 14-20	1.5e+00 [3.0e-02;1.4e+02]
maternalage factor 3-56	1.0e-02 [0.0e+00;2.3e-01]
nr siblings	0.0e+00 [0.0e+00;0.0e+00]
older siblings 1	0.0e+00 [0.0e+00;2.0e-02]
older siblings 2	1.0e+00 [2.0e-02;5.3e+01]
older siblings 3	2.0e+06 [1.7e+04;2.8e+08]
older siblings 4	1.5e+07 [6.0e+04;3.0e+09]
older siblings 5+	1.3e+11 [1.2e+08;1.5e+14]
paternal loss 0-1	1.4e+00 [0.0e+00;2.2e+04]
paternal loss 1-5	0.0e+00 [0.0e+00;5.7e+01]
paternal loss 10-15	3.5e+01 [3.0e-02;7.1e+04]
paternal loss 15-20	3.0e-02 [0.0e+00;1.1e+01]
paternal loss 20-25	7.0e-02 [0.0e+00;3.9e+01]
paternal loss 25-30	2.4e+00 [2.0e-02;5.9e+02]
paternal loss 30-35	7.5e+00 [4.0e-02;2.9e+03]
paternal loss 35-40	1.0e-02 [0.0e+00;2.7e+00]
paternal loss 40-45	2.7e-01 [0.0e+00;7.8e+03]
paternal loss 5-10	5.2e+01 [2.0e-02;3.5e+05]
paternalage	0.0e+00 [0.0e+00;0.0e+00]
paternalage mean	2.8e+04 [4.9e+02;1.4e+06]

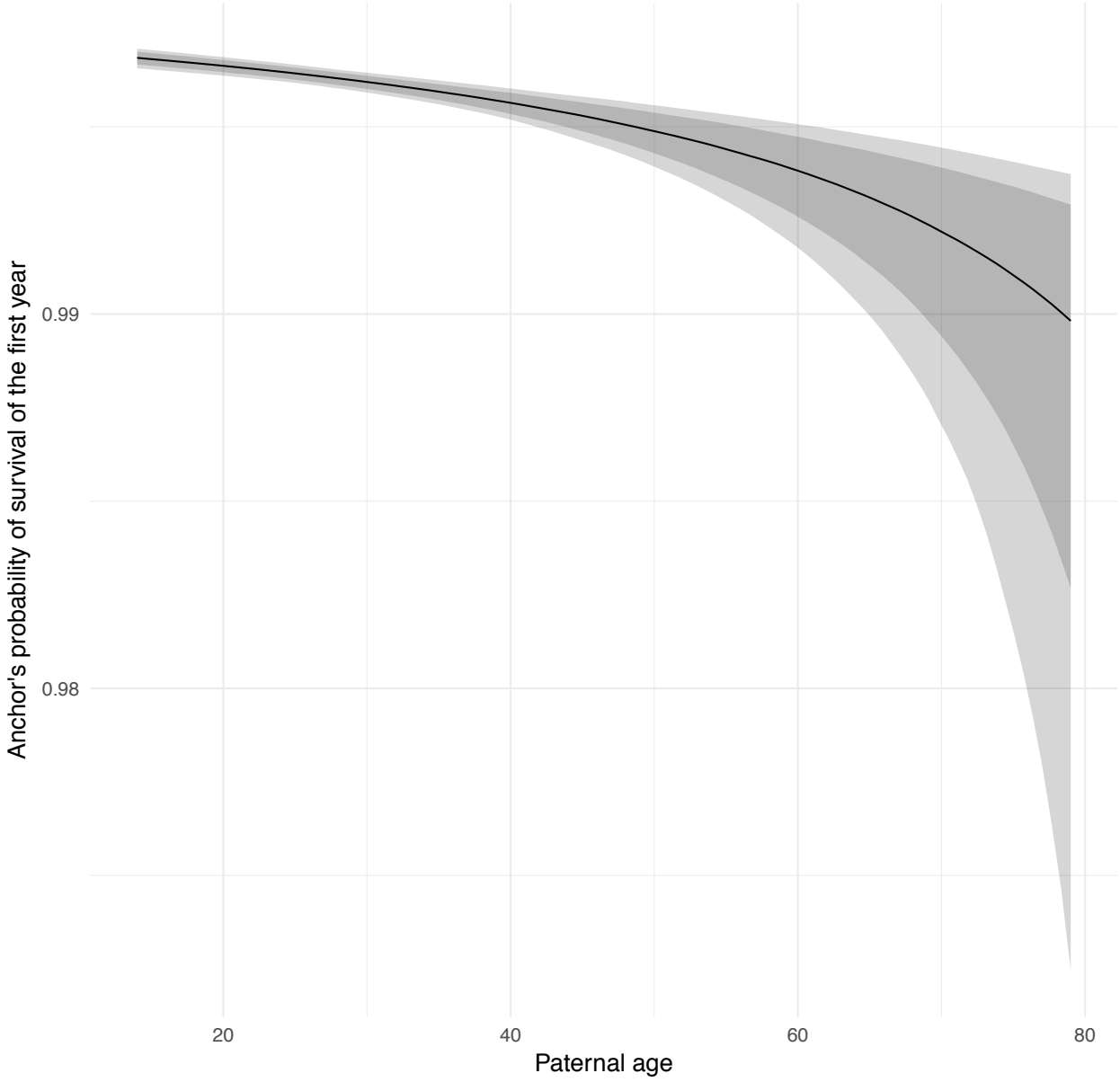
5.6.5 Paternal age effect

This is the effect of 10 years of paternal age within families on probability of survival of the first year, combined over the hurdle and Zero-truncated Poisson component, expressed as a change in percentage `((predicted value at t + 10y)/(predicted value at t)) - 1`.

effect	median_estimate	ci_95	ci_80
percentage change	-0.05	[-0.06;-0.03]	[-0.06;-0.03]

5.6.5.1 Marginal effect plot

Paternal age effect on probability of survival of the first year The shaded areas show the 95% and 80% credibility intervals for the reference individuals and include uncertainty related to covariate effect sizes.



6 Robustness analyses documentation

All of the following models are the same as our main model m3, except for the noted changes to test robustness.

6.1 Table

Estimates of the effect of 10 years paternal age on number of children within families (comparing siblings) are given in percentage change. 95% credibility interval are given in brackets. Click the model name to be taken to further details on the supplementary website.

Population	Model	Estimate
20th-century Sweden	m3 model with 95% CI	-5.15 [-5.94;-4.36]
Historical Sweden	m3 model with 95% CI	-7.29 [-13.40;-1.07]
Krummhörn	m3 model with 95% CI	-8.41 [-24.83;12.03]
Québec	m3 model with 95% CI	-3.00 [-6.08;0.24]
20th-century Sweden	r1 relaxed exclusion criteria	-4.96 [-7.67;-2.18]
Historical Sweden	r1 relaxed exclusion criteria	-17.32 [-22.85;-11.80]
Krummhörn	r1 relaxed exclusion criteria	-7.11 [-23.86; 6.73]
Québec	r1 relaxed exclusion criteria	-1.67 [-4.39;0.62]
20th-century Sweden	r2 few controls	-3.20 [-4.37;-2.00]
Historical Sweden	r2 few controls	-20.42 [-23.68;-17.37]
Krummhörn	r2 few controls	-16.37 [-22.17;-10.38]
Québec	r2 few controls	-6.62 [-7.96;-5.24]
20th-century Sweden	r3 birth order continuous	-2.08 [-2.87;-1.29]
Historical Sweden	r3 birth order continuous	-6.02 [-12.94; 1.64]
Krummhörn	r3 birth order continuous	-23.54 [-37.92;-5.40]
Québec	r3 birth order continuous	-1.68 [-6.18;3.55]
20th-century Sweden	r4 control dependent sibs	-3.97 [-5.37;-2.56]
Historical Sweden	r4 control dependent sibs	-2.58 [-6.01;1.07]
Krummhörn	r4 control dependent sibs	0.68 [-8.48;10.93]
Québec	r4 control dependent sibs	0.53 [-1.33;2.46]
20th-century Sweden	r5 birth order interact siblings	-4.52 [-7.05;-1.85]
Historical Sweden	r5 birth order interact siblings	-6.80 [-12.77;-0.16]
Krummhörn	r5 birth order interact siblings	-10.72 [-28.05; 9.92]
Québec	r5 birth order interact siblings	-0.93 [-4.21;2.60]
20th-century Sweden	r6 no birth order control	-3.70 [-5.13;-2.29]
Historical Sweden	r6 no birth order control	-2.34 [-5.7802;1.3217]
Krummhörn	r6 no birth order control	1.05 [-8.48;11.09]
Québec	r6 no birth order control	0.28 [-1.73;2.28]
20th-century Sweden	r7 less parental loss control	-4.84 [-7.21;-2.19]
Historical Sweden	r7 less parental loss control	-24.13 [-30.14;-18.17]
Krummhörn	r7 less parental loss control	-18.19 [-34.37;-0.24]
Québec	r7 less parental loss control	-7.42 [-10.47;-4.22]
20th-century Sweden	r8 adjust for first born adult	-4.80 [-7.20;-2.42]
Historical Sweden	r8 adjust for first born adult	-11.03 [-16.11;-5.49]
Krummhörn	r8 adjust for first born adult	-7.93 [-21.26; 8.17]
Québec	r8 adjust for first born adult	-8.63 [-12.34;-4.75]
20th-century Sweden	r9 continuous byear adjustment	-4.60 [-7.24;-1.98]
Historical Sweden	r9 continuous byear adjustment	-7.04 [-12.40;-1.31]
Krummhörn	r9 continuous byear adjustment	-6.35 [-21.47;11.08]
Québec	r9 continuous byear adjustment	-3.86 [-7.35;-0.26]
20th-century Sweden	r10 add random slope	-4.57 [-7.15;-2.11]

Population	Model	Estimate
Historical Sweden	r10 add random slope	-23.70 [-29.41;-17.92]
Krummhörn	r10 add random slope	-8.56 [-25.31;11.57]
Québec	r10 add random slope	-5.17 [-7.98;-1.30]
20th-century Sweden	r11 separate random effects for parents	-4.61 [-7.14;-1.90]
Historical Sweden	r11 separate random effects for parents	-7.30 [-13.43;-0.80]
Krummhörn	r11 separate random effects for parents	-8.41 [-25.14;11.23]
Québec	r11 separate random effects for parents	-2.68 [-5.95;0.52]
20th-century Sweden	r12 sex moderation	-3.94 [-6.61;-1.18]
Historical Sweden	r12 sex moderation	-7.88 [-13.91;-1.34]
Krummhörn	r12 sex moderation	-9.05 [-25.49;11.82]
Québec	r12 sex moderation	-3.24 [-6.43;0.31]
20th-century Sweden	r13 control paternal afb	-4.15 [-6.65;-1.65]
Historical Sweden	r13 control paternal afb	-8.34 [-14.90;-2.41]
Krummhörn	r13 control paternal afb	-8.15 [-24.89;12.04]
Québec	r13 control paternal afb	-2.19 [-5.22986;1.19428]
20th-century Sweden	r14 compare lfe	
Historical Sweden	r14 compare lfe	
Krummhörn	r14 compare lfe	
Québec	r14 compare lfe	
Historical Sweden	r15 region moderator parish ranef	-6.89 [-12.58;-1.25]
Krummhörn	r15 region moderator parish ranef	-8.31 [-24.61;10.99]
Québec	r15 region moderator parish ranef	-2.17 [-5.16;0.79]
Historical Sweden	r16 restrict to skelleftea	-10.22 [-22.12; 4.20]
20th-century Sweden	r17 simulate downs	-4.09 [-6.66;-1.24]
Historical Sweden	r17 simulate downs	0.41 [-6.23;7.48]
Krummhörn	r17 simulate downs	2.67 [-15.43;25.56]
Québec	r17 simulate downs	0.96 [-2.48;4.45]
20th-century Sweden	r18 hurdle poisson	-3.14 [-5.48;-0.71]
Historical Sweden	r18 hurdle poisson	-8.54 [-12.54;-4.62]
Krummhörn	r18 control paternal afb	-7.10 [-24.21;13.41]
Krummhörn	r18 hurdle poisson	-2.09 [-13.04;10.40]
Québec	r18 hurdle poisson	-2.61 [-5.0581;0.0065]
20th-century Sweden	r19 normal distribution	-3.73 [-6.14;-1.40]
Historical Sweden	r19 normal distribution	-6.48 [-12.63;-0.85]
Krummhörn	r19 normal distribution	-7.69 [-21.51; 8.34]
Québec	r19 normal distribution	-4.32 [-7.39;-1.11]
20th-century Sweden	r20 no maternalage control	-5.28 [-7.65;-2.98]
Historical Sweden	r20 no maternalage control	-5.55 [-11.28; 0.50]
Krummhörn	r20 no maternalage control	-13.61 [-29.41; 3.73]
Québec	r20 no maternalage control	-3.49 [-6.27;-0.66]
20th-century Sweden	r21 continuous maternalage	-0.94 [-1.74;-0.13]
Historical Sweden	r21 continuous maternalage	-8.94 [-15.31;-2.32]
Krummhörn	r21 continuous maternalage	-10.86 [-28.22; 8.72]
Québec	r21 continuous maternalage	-1.75 [-5.37;1.84]
Historical Sweden	r22 relaxed exclusion censoring	-25.50 [-29.91;-21.11]
Krummhörn	r22 relaxed exclusion censoring	-6.76 [-23.07; 6.41]
Québec	r22 relaxed exclusion censoring	-1.74 [-4.43;0.41]
20th-century Sweden	r23 student cauchy priors	-4.60 [-7.26;-2.03]
Historical Sweden	r23 student cauchy priors	-7.27 [-13.28;-1.01]
Krummhörn	r23 student cauchy priors	-8.39 [-25.46;10.78]
Québec	r23 student cauchy priors	-3.04 [-6.17;0.21]
20th-century Sweden	r24 uniform priors	-4.71 [-7.13;-2.07]

Population	Model	Estimate
Historical Sweden	r24 uniform priors	-7.25 [-13.17;-1.03]
Krummhörn	r24 uniform priors	-8.90 [-26.92;11.29]
Québec	r24 uniform priors	-3.00 [-6.25;0.18]
Historical Sweden	r25 migration status	-10.44 [-16.68;-3.48]
Krummhörn	r25 migration status	-1.98 [-28.80;31.32]
Québec	r25 migration status	-2.98 [-6.22;0.38]
20th-century Sweden	r26 separate parental age contributions	-2.22 [-3.88;-0.44]

6.2 Model descriptions

6.2.1 *r1*: Relaxed exclusion criteria

For the three historical populations, we imposed quite stringent exclusion criteria to ensure sufficient data quality for our intended analysis. This was not necessary for the modern Swedish data, because there were no exclusion criteria to relax.

6.2.2 *r2*: Fewer covariates

Adding covariates increases the complexity of the model and makes it harder to interpret. We chose to adjust for many potential confounds because we are interested in causal isolation of the paternal age effect. Here we show what happens when only birth cohort and average paternal age in the family are adjusted for.

6.2.3 *r3*: Continuous birth order control

We chose to control for birth order/number of older siblings as a categorical variable, lumping all those who had more than 5 in the category 5+. Because a continuous covariate is also plausible, we tested this alternative model as well.

6.2.4 *r4*: Control number of dependent siblings

Birth order is usually used as a proxy variable for parental investment, the assumption being that older siblings require parental attention. However, there are are reasons to doubt this, as fully-grown siblings probably do not compete for the same resources. To compute a clearer proxy variable of competing siblings, we computed and adjusted for the number of siblings who were alive and younger than five at the time of birth of the anchor child.

6.2.5 *r5*: Birth order interacted with number of siblings

Plausibly, being first-born has a different effect, when one is an only child as opposed to having two siblings, etc. Here, we allow for such an interaction effect.

6.2.6 *r6*: No birth order control

Paternal age and birth order are highly collinear with each other and with maternal age. Therefore, the choice to include this predictor widens standard errors for each predictor and may be disputed. Here we show what happens when we simply omit the birth order control.

6.2.7 *r7*: Less control for parental loss

We adjusted for parental loss very stringently, including covariates for parental loss up to age 45. Here we show what happens, when we only control for parental loss in the first, and the first five years of life.

6.2.8 *r8*: Adjust for being first-/last-born adult son

Inheritance is linked to birth order and being male in several of the historical populations. Here, we adjust for the anchor being the first or last born adult son in a family. This implies that we control for our outcome to a certain extent, as "adult sons" cannot have died before adulthood, but a paternal age effect on mortality could still be detected for siblings other than the first- and last-born adults.

6.2.9 *r9*: Continuous birth year adjustment

In our main model, we control for birth cohort in 5-year-bins (lumping small bins). We chose to do so, because nonlinear and even sharply spiking effects of birth cohort are plausible (due to e.g. epidemics). This decision may be disputed, as it summarises 5-year-bins. Here, we instead allow for a thin-splate spline on the continuous birth year variable. This allows for smooth nonlinear (but not spiking) birth cohort effects.

6.2.10 *r10*: Group-level slope added

Paternal age effects may vary between different families. Although we did not explore between-family moderators of paternal age effects in our study, we tested whether modelling an additional group-level slope for paternal age differences within the family, would change the results by allowing for shrinkage and to examine the amount of inter-family differences to be explained for potential future moderator analysis.

6.2.11 *r11*: Separate group-level effects for each parent

Most anchors in our sample are full biological siblings and especially in the historical populations, divorce and remarriage was rare. Therefore, we chose to include only one group-level effect, for the parent couple (i.e. one group-level effect per father-mother-dyad). Including one intercept per parent is potentially a better way to adjust for genetic propensities inherited from either parent and allows estimating this propensity also from half-siblings, while half-sibling relationships were ignored in our main models. This comes at the cost of modelling complexity.

6.2.12 *r12*: Sex moderation

It need not be the case that paternal age has the same effect on male and female children. For example, male children inherit only the small Y chromosome from the father, but female children inherit the larger X chromosome, so that paternal age predicts X-chromosomal de novo mutations in females but not in males (Francioli et al., 2016). At the same time, the autism literature suggests that males are less robust to heritable and de novo autism risk variants and that these effects are not simply due to having only one X chromosome (Werling & Geschwind, 2015). Here we let a dummy variable for being male moderate the paternal age effect.

6.2.13 *r13*: Control paternal age at first birth

We already control for the average paternal age at which the children in a family were born. The mean is a more complete summary of the reproductive timing of the father than the age at first birth. However, far more literature has examined age at first birth and it has the advantage of never being censored (although we

of course try to rule out censoring by choosing appropriate subsets). Therefore, we added age at first birth as a covariate in this model.

6.2.14 *r14*: Compare lfe

Most of the previous literature has not used multilevel modelling, but linear group fixed effects (essentially dummy variables on the many thousands of families in the model). We believe our multilevel modelling approach has the advantage of allowing us to examine the effect of including predictors at the level of the family in the same model.

This allows us to
a) appropriately model a zero-inflated outcome such as number of children including those who died young (we're not aware of a linear group fixed effect approach that handles hurdle or zero-inflated models)
b) examine group-level slopes for paternal age and potentially to examine moderators at the level of the family (though we did not do this)
c) explicitly model confounders at the level of the family (e.g. number of siblings).

Nevertheless, the prevalence of this approach in the literature mandates that we show how our approach compares. We fit this model using the R package "lfe" and the function felm. All covariates that were not estimable in principle were removed (i.e. number of siblings, paternalage.mean).

Because we cannot extract an effect size comparable to the other models from these models, these results are viewable only online.

6.2.15 *r15*: Using a moderator by region, group-level effects by parish

In this model we attempted allow for regional variation in paternal age effects and attempted to better control residual variation. Our approach was two-fold: to moderate paternal age by region and to add a random effect for the church parish in which the individual was born. However, for the modern Swedish data, we had no geographic data and no regional information, so this model was not fit.

6.2.16 *r16*: Restrict to Skellefteå

Only in the DDB (historical Swedish data), parishes in some of the regions were still unlinked. This means that individuals could occur in more than one parish and not be linked. However, the region of Skellefteå was fully linked. Here, we test what happens when we restrict our dataset to Skellefteå.

6.2.17 *r17*: Simulating Down syndrome cases

1. We assume that 4 in 1000 births are children with Down syndrome (four times the actual rate).
2. We randomly excluded 33% of all children who had a mother older than 40 and had no children (many times the actual rate at that age).

6.2.18 *r18*: Reversing hurdle_poisson and poisson

To make models computationally feasible and because early mortality was negligible, we fit the very large modern Swedish dataset with a `poisson()` family distribution. All historical datasets had high early mortality, so we thought a `hurdle_poisson()` was more appropriate. Here, we show what happens when we reverse this. The `hurdle_poisson()` model can be fit to the modern Swedish data here, because we only use a subset.

6.2.19 *r19*: Normal distribution

Previous analysts sometimes decided to use the normal distribution to predict (potentially zero-inflated) count data. Here, we refit our models using a normal distribution for the outcome. We show that estimates for the paternal age effect can be estimated to have a substantially different magnitude, because of this, but did not change direction.

6.2.20 *r20*: No adjustment for maternal age

In this model, we test what happens when we do not adjust for maternal age, because it is highly collinear with paternal age.

6.2.21 *r21*: Continuous adjustment for maternal age

In this model, we adjust for maternal age using a continuous variable instead of three bins. This does not allow for nonlinear effects, but also does not aggregate the predictor. We cannot compare full siblings, test the effects of maternal and paternal age and adjust for average maternal and paternal age in the family (because the predictors are redundant), so that it is not perfectly possible to disentangle the contribution of maternal and paternal age and compare full siblings.

6.2.22 *r22*: Relaxed exclusion and censoring criteria

Like *r1*, but we use a 30-years-later cutoff year for our birth cohorts, relaxing our censoring requirements.

6.2.23 *r23*: Student's t and half-Cauchy priors

To demonstrate the robustness of our prior choice we use Student's t priors (fatter tails than normal priors) for our population-level effects and a half-Cauchy prior for our group-level effect for the family.

6.2.24 *r24*: Improper flat priors

To demonstrate the robustness of our prior choice we use improper flat priors. These priors should make the model's results comparable to a frequentist maximum likelihood approach.

6.2.25 *r25*: Adjust for migration status

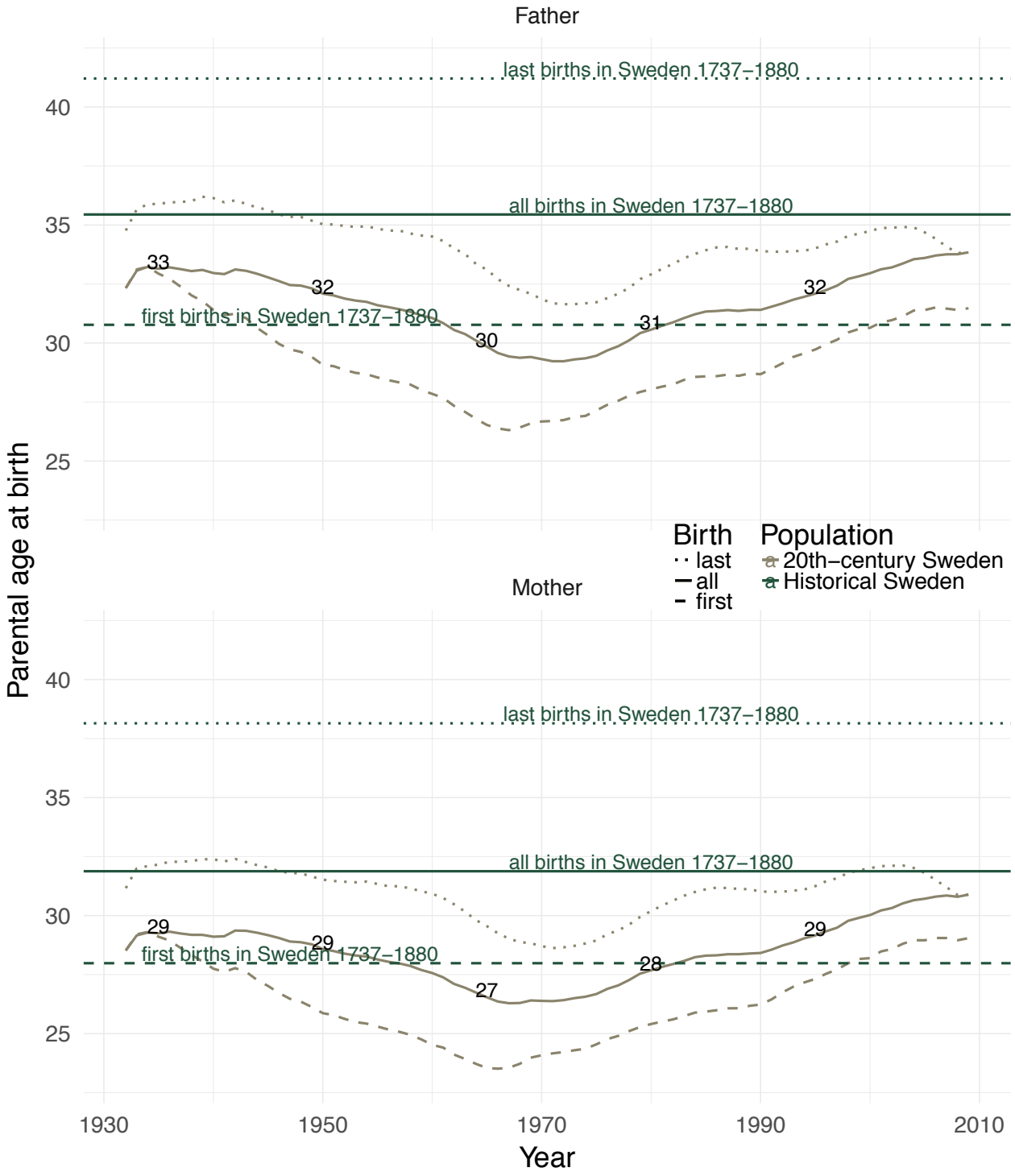
In the three historical populations, records were kept in the parish. Although records were linked between parishes in all populations, except three out of four provinces in historical Sweden, migration might sometimes lead to censoring of records. Adjusting for migration may however constitute a partial adjustment for the outcome, as lower offspring fitness might make them more likely to migrate. Hence, we show the results of doing so as a robustness analysis. In all analyses, we adjusted for a "migrated"-dummy variable. Migration was differently defined depending on the population. In Québec, we had flags denoting immigrants and emigrants. Few immigrants were included in our analyses anyway, as we needed parental information for our analyses. Emigrants were people who left Québec. In historical Sweden, migration was logged as migration from the parish of birth. In the Krummhörn, we set migrated to true, when the parish of death/burial differed from the parish of birth/baptism.
No migration information was available in 20th-century Sweden, but records there weren't kept in parishes, so this should not pose a problem.

6.2.26 *r26*: Separate parental age contributions

In this model, we adjust for maternal age using a continuous variable. We also adjust for a dummy variable for teenage motherhood, to account for the nonlinearity of the maternal age effect. Moreover, we use separate random intercepts for mothers and fathers and adjust for the mother's mean age at birth and the father's mean age at birth. This model only converges in the 20th-century Sweden data, because there are sufficient numbers of divorces and remarriages and enough data to separate the parents' contributions.

7 Reproductive timing in Sweden

Reproductive timing data showed that average parental ages at birth decreased in 20th-century Sweden until ca. 1970 and increased thereafter. Average contemporary parental ages are still lower than in any of the three historical populations. Ages at first birth in the early periods and ages at last birth in the late periods are censored and hence biased towards the age at all births, which is itself unbiased.



Appendix C.

Manuscript 3 (Using 26 thousand diary entries to show ovulatory changes in sexual desire and behaviour)

Using 26 thousand diary entries to show ovulatory changes in sexual desire and behaviour

Authors: Ruben C. Arslan[1,2], Katharina M. Schilling[1], Tanja M. Gerlach[1,2†], Lars Penke[1,2†]

[†] LP and TMG share the last authorship.

Affiliations:

1. Biological Personality Psychology,
   Georg Elias Müller Institute of Psychology,
   University of Goettingen,
   37073 Göttingen, Germany
2. Leibniz ScienceCampus Primate Cognition,
   37073 Göttingen, Germany

Author Note: This research has been previously presented at the Human Behavior and Evolution Conference (2017), at the European Human Behaviour and Evolution Association (2016), and the Congress of the German Society of Psychology (2016). Preliminary results from this study formed the basis of KMS' master thesis. The authors have no conflicts of interest.

Study documentation: https://osf.io/kd26j/

Supplementary website: https://rubenarslan.github.io/ovulatory_shifts/

Abstract

Previous research reported ovulatory changes in women's appearance, mate preferences, extra- and in-pair sexual desire and behaviour, but has been criticised for small sample sizes, inappropriate designs, and undisclosed flexibility in analyses. In the present study, we sought to address these criticisms by preregistering our hypotheses and analysis plan and by collecting a large diary sample. We gathered over 26 thousand usable online self-reports in a diary format from 1043 women, of which 421 were naturally cycling. We inferred the fertile period from menstrual onset reports. We used hormonal contraceptive users as a quasi-control group, as they experience menstruation, but not ovulation. We probed our results for robustness to different approaches (including different fertility estimates, different exclusion criteria, adjusting for potential confounds, moderation by methodological factors). We found robust evidence supporting previously reported ovulatory increases in extra-pair desire and behaviour, in-pair desire, and self-perceived desirability, as well as no unexpected associations. Yet, we did not find predicted effects on partner mate retention behaviour, clothing choices, or narcissism. Contrary to some of the earlier literature, partners' sexual attractiveness did not moderate the cycle shifts. Taken together, the replicability of the existing literature on ovulatory changes was mixed. We conclude with simulation-based recommendations for reading the past literature and for designing future large-scale preregistered within-subject studies to understand ovulatory cycle changes and the effects of hormonal contraception. Interindividual differences in the size of ovulatory changes emerge as an important area for further study.

Keywords: ovulatory cycle shifts, sexual desire, diary study, hormonal contraception, evolutionary psychology

1 Introduction

1.1 Theoretical Background

Personality, behaviour, sexual desire, attractiveness, mate preferences and mate choices vary between and within persons (Fleeson, 2001, 2004; Gerlach, Arslan, Schultze, Reinhard, & Penke, in press). While copious research has identified antecedents of interindividual variation (Zietsch, Lee, Sherlock, & Jern, 2015), it is still often viewed as mere chance fluctuation or response to situational demands. Systematic endogenous causes of intraindividual variation are worthy of further study.

In the evolutionary psychology literature, the menstrual cycle has been suggested as one such influence on psychological state fluctuations in women (Gangestad & Thornhill, 2008). Menstrual cycle changes in attractiveness, mate preferences, and sexual desire, as well as men's reactions to those changes have been interpreted as evidence for adaptations formed by sexual selection and sexually antagonistic coevolution, i.e. arms races between the sexes. However, to this day debate continues over the existence and extent of such changes (W. Wood, Kressel, Joshi, & Louie, 2014).

In this paper, we have the twin goals of reviewing methodological problems with commonly used approaches and addressing them in a high-powered, preregistered replication study. Because our study was preregistered in March 2014, the introduction of this manuscript reflects our reading of the literature at that point in time. We review recent theoretical and empirical developments in the discussion.

1.1.1 Do human females show oestrus?

Human women do not develop garish sexual swellings or other prominent changes around ovulation, unlike their closest cousins, the chimpanzees (Deschner, Heistermann, Hodges, & Boesch, 2003). Moreover, human women and several other primates exhibit *extended sexuality,* that is they have sex outside the fertile window, not just during a period of oestrus or *heat* (Dixson, 2012).

However, other, less conspicuous endocrine, behavioural, physiological and psychological changes happen over the course of the menstrual cycle and some peak when women are fertile (Gangestad & Simpson, 2000; Haselton & Gildersleeve, 2016). This led (Gangestad & Thornhill, 2008) to argue that the differentiation of functional and physiological aspects of fertile phase sexuality merits being called oestrus.

1.1.2 The good genes ovulatory shift hypothesis

The ovulatory shift hypothesis posits that women's mate preferences and choices vary with their fertility status. It is a central functional differentiation predicted under the human oestrus perspective (Gangestad & Thornhill, 2008). According to this theory, women would optimise their reproductive potential by choosing to be with partners who will invest in offspring during non-fertile times and choosing, if necessary, other, extra-pair, males with *good genes* to provide their offspring's genes, i.e. to have sex with during the fertile phase. To differentiate this theoretically predicted *ovulatory shift in mate preferences to obtain good genes, potentially from extra-pair copulations* (Pillsworth & Haselton, 2006a) from simpler, generalized increases in sexual drive or

libido in the fertile phase, we will call this theory *good genes ovulatory shift hypothesis* (GGOSH).

The concept of *good genes* is meant to index genetic qualities that women should want their offspring to inherit. The concept includes dyadic genetic fit (e.g. good immunocompetence genes), genetic fit to the current environment, and few harmful mutations. It has no direct correspondence in the evolutionary genetic literature and some purported indicators of *good genes* are controversial (Arslan & Penke, 2015).

Several male characteristics have been argued to indicate *good genes.* Cycle studies have then reported fertile phase increases in preferences for these traits, which include masculinity, low fluctuating asymmetry (Scheib, Gangestad, & Thornhill, 1999), and various measures of attractiveness (Gildersleeve, Haselton, & Fales, 2014a; Haselton & Gangestad, 2006; Larson, Haselton, Gildersleeve, & Pillsworth, 2013; Pillsworth & Haselton, 2006b). In laboratory studies, fertile phase shifts towards preferences for male stimuli with such characteristics (photos, videos, voice samples), have been cited as support for GGOSH (Gildersleeve et al., 2014a).

1.1.3 Rationale for the present study

In our study, we sought to replicate and extend previous results from field studies of naturally cycling women commonly cited as evidence of a differentiation of fertile phase sexuality. These field studies reported evidence for changes in female sexual interests and appearance across the cycle. Central results in these studies served as the rationale for the preregistration of our study.

1.1.3.1 Extra-pair desire and behaviour

(Gangestad, Thornhill, & Garver, 2002) asked 51 naturally cycling women (i.e. not using hormonal contraceptives) to report their sexual interests and fantasies once in the fertile and once in the non-fertile phase. Women reported substantially greater attraction to and fantasies about men other than primary partners when fertile.

In a sample of 54 couples and using the same study design, (Gangestad, Thornhill, & Garver-Apgar, 2005) additionally reported support for a predicted moderator effect. Women showed stronger fertile phase increases in attraction to other men if paired with relatively asymmetrical primary partners. In a diary study, (Haselton & Gangestad, 2006) asked 38 naturally cycling women to provide daily reports of sexual interest and feelings for 35 days. Women reported that they were more attracted to and flirted more often with men other than primary partners on higher fertility days, if their partner's sexual attractiveness was low.

1.1.3.2 In-pair desire and behaviour

According to the ovulatory shift hypothesis, women whose long-term partners display indicators of "good genes" do not benefit from engaging in what (Pillsworth & Haselton, 2006a) call a dual-mating strategy. The authors predicted such women should instead experience ovulatory increases in in-pair desire. Findings were mixed, with some showed the predicted moderated shifts (Gangestad et al., 2005; Pillsworth, Haselton, & Buss, 2004) others did not (Gangestad et al., 2002; Pillsworth & Haselton, 2006b). (Gangestad et al., 2002) found that women did not experience significantly

higher levels of overall sexual desire when fertile, but tended to initiate and have more sex with their partners as ovulation neared.

1.1.3.3 Male mate retention

Because female extra-pair sex might lead her primary partner to involuntarily invest parental care and resources into offspring sired by an extra-pair mate, counter-adaptations to the aforementioned shifts were predicted (Pillsworth & Haselton, 2006a). (Gangestad et al., 2002) correspondingly found that *prohibitive* (i.e. jealousy) and *persuasive* (i.e. affection) male partners' mate retention tactics increased during the fertile phase. (Haselton & Gangestad, 2006) replicated these results. These tactics were exhibited primarily by partners of women who perceived their partners to be low in sexual attractiveness relative to investment attractiveness.

1.1.3.4 Self-perceived desirability and clothing choices

Although obvious outward signals of fertility are absent in humans, some studies report evidence of subtle ovulatory cues in human females and conclude that ovulation may not be perfectly concealed. (Haselton & Gangestad, 2006) reported that women perceived themselves to be more attractive when fertile. Haselton et al. (2007) further predicted and found fertile phase increases in grooming and attractive clothing choices in a sample of 30 partnered women who were photographed at high and low fertility. (Schwarz & Hassebrauck, 2008) replicated and extended this study. In a sample of 40 women who completed a daily questionnaire over 31 days, participants rated their perceived attractiveness, and their clothing style on the dimensions "figure-hugging", "sexy", and "permissive". They were also instructed to take one photo of themselves

each day. Men then rated these photos for clothing style and physical attractiveness. Women perceived themselves and were perceived by men to be dressed more provocatively on their fertile days. In another replication, using 88 women tested twice, (Durante, Li, & Haselton, 2008) reported evidence that women prefer clothing that is more revealing and sexy during the fertile phase, as shown in full-body photographs and drawn illustrations of what they would wear to a hypothetical social event that evening.

1.1.3.5 Intrasexual competitiveness

Durante et al. (2008) interpreted their results discussed above as evidence of increased intrasexual competitiveness, i.e. women altering their physical appearance to enhance their ability to compete with other women. We speculated that, if intrasexual competitiveness during the fertile phase were increased, we might detect this in narcissistic personality states, as conceptualized in the two-dimensional narcissistic admiration and rivalry concept (NARC(Back et al., 2013). Narcissistic admiration is thought to be linked to the desire to attain social status, and evoke social interest. Narcissistic rivalry is thought to be linked to motivations to defend one's social status against others. In the context of our study, to test the prediction of increased intrasexual competitiveness in the fertile phase (Durante et al., 2008) in a novel way, we reformulated narcissistic state items for both NARC dimensions to refer to other women only.

1.2 Methodological issues

The psychological literature on ovulatory changes has been criticised and hotly debated. Two meta-analyses based on overlapping data both concluded that publication bias afflicts research on ovulatory shifts in mate preferences, as may be the case for most of the scientific literature (Fanelli, 2011; Ferguson & Brannick, 2012). However, one team of investigators (Gildersleeve et al., 2014a) concluded that all evidence taken together suggested replicable shifts in mate preferences, even after including studies freed from the file drawer and adjusting for bias. Another team (W. Wood et al., 2014) concluded further bias and methodological artefacts implied that any non-negligible effects were, in fact, overestimated. Our study focuses on different outcomes than these meta-analyses, but many of the criticisms and problems pertain to the designs commonly used to study ovulatory change, irrespective of outcomes and research questions. Thus they also influenced our approach. In the following, we summarise several methodological issues brought to the fore by this debate.

1.2.1 Researcher degrees of freedom can lead to false positives

Many psychological studies do not replicate in exact replications (Open Science Collaboration, 2015). Potential sources of bias are *researcher degrees of freedom* in specifying hypothesis, methodology, and statistical approach after seeing the data. Journals and researchers tend to preferentially publish and cite significant counter-intuitive results, leading to warped incentives (Simmons, Nelson, & Simonsohn, 2011).

Recent debate in the menstrual cycle literature has specifically highlighted flexibility in the definition of the fertile window, but more general problems such as reporting only

significantly associated measures and stopping data collection conditional on significance could also affect the literature. As surveys of psychological researchers show that some research practices now deemed questionable were widespread (John, Loewenstein, & Prelec, 2012) and meta-analyses show publication bias. Both sides in the ovulatory cycle debate acknowledge bias (Gangestad, 2016; Harris, Pashler, & Mickes, 2014; W. Wood et al., 2014) but do not agree on whether and how it can be adjusted for (Gildersleeve, Haselton, & Fales, 2014b; Harris et al., 2014) in order to obtain trustworthy bias-corrected estimates (Inzlicht, Gervais, & Berkman, 2015; van Elk et al., 2015). The debate surrounding this has at times turned vitriolic, because the often used term *p-hacking* has connotations of intentional mischief, but it is clear from simulations (Smaldino & McElreath, 2016) and intuition (Gelman & Loken, 2014) that flexibility will lead to bias even without ill intentions, as long as odds of publication and tenure can hinge on whether results turn statistically significant. Ultimately, although methods such as the *p-curve* (Gildersleeve et al., 2014b) can offer suggestive evidence of replicability, the true tests of replicability are *preregistered* replication studies in which hypotheses, methods and statistical approach are fixed before the data are collected, preventing *researcher degrees of freedom* from skewing results.

1.2.2 Estimating the fertile window

There is wide variability in the approaches used to estimate women's fertile windows. (Gildersleeve et al., 2014a) reviewed these approaches and problems associated with them. (Gangestad et al., 2016) recommend that researchers abandon windows altogether and instead estimate continuous probabilities of being fertile.

Flawed recall of the last menstrual onset, accuracy being as low as 57% (Wegienka & Baird, 2005), remains a problem. Moreover, menstrual cycle lengths vary within person, so that recalled average cycle length correlates only ~.5 with the length of individual cycles (Blake, Dixson, O'Dean, & Denson, 2016; Gangestad et al., 2016). Because the follicular phase leading up to ovulation is more variable than the luteal phase (Fehring, Schneider, & Raviele, 2006), the more convenient method (forward counting from the last menstrual onset) is also more imprecise (Gangestad et al., 2016). Backward counting to ovulation from the next menstrual onset should hence be more accurate, with a validity for estimated fertility as high as ~.7 (Gangestad et al., 2016). (Blake et al., 2016) report much lower validities, using luteinising hormone (LH) surges as the criterion in a small sample of 140 women, but re-analyses of their data using a hedged fertile window estimate, as in Gangestad et al. (2016), show comparable validities.

For researchers, backward counting has the added benefit that women who count days as part of their contraception regiment cannot do it prospectively, perhaps reducing awareness and thus demand characteristics. Still, counting-based estimates of conception probability derive from forward-counted actuarial values which are then reversed (Gangestad et al., 2016), ideally actuarial estimates would be backward-counted too.

1.2.3 Between-subject designs to study a within-subject process

Many past studies have used between-subject designs to study a within-subject process, ovulation (Gangestad et al., 2016). Even when sample sizes are large, selection bias could confound any identified effects. One possible scenario could be that a common cause, for instance genetic makeup or a disease, makes women anovulatory and lowers their sexual desire. This could lead researchers to mistake a between-subject difference for an ovulatory change. Another potential problem might be that increased social activity during the fertile phase (Haselton & Gangestad, 2006) could make fertile women less likely to participate in a survey study, biasing the sample towards women who experience smaller changes. Further, cross-sectional designs can never reliably measure individual differences in the size of ovulatory changes. They may also lead to the use of outcome measures that measure a trait component, but not a state component, reliably. This can be avoided by using established measures tested on within-subject data. Indeed, many of the above problems are minor and could potentially be avoided or adjusted for, but given that within-subject studies do not have these problems and are no longer hard to implement, they seem the superior option. Most crucially however, typical between-subject studies have far too low statistical power at typical samples sizes, as shown by (Gangestad et al., 2016).

1.2.4 Lack of power or implausible effect size expectations

The average menstrual cycle study is underpowered to detect anything but very large changes (Gangestad et al., 2016). At the same time, most researchers seem to agree that ovulatory changes are, if anything, subtle. In this situation, many plausible and interesting effect sizes will be missed, and reported effects will tend to be

overestimates. If we desire theoretical progress, we need to narrow down effect sizes to disambiguate between theories that predict no, minimal, small, medium, or large ovulatory changes in certain outcomes. Thus, the literature would benefit from narrower confidence intervals to resolve theoretical debates over evolutionary function. Even for larger effects, typical cycle studies are underpowered, because of the combination of suboptimal design aspects and small sample size (median $N = 48$ in Gildersleeve, Haselton, and Fales, 2014, mean $N = 49$ in the studies we sought to replicate). For between-subject studies planning to achieve 80% power to detect a Cohen's d of 0.4 with a backward-counted conception probability estimate, Gangestad et al. (2016) recommend a sample size of 1,143.

1.2.5 No differentiation of women by reproductive intentions and contraception method

(W. Wood et al., 2014) pointed out that the most uniquely human aspect of menstrual cycles may be women's exertion of control over their cycle and fertility to adapt to cultural, societal and their own needs. Although they provide no specific recommendations how this should change research practices, we note that most studies do not report differentiating between naturally cycling women who use barrier methods, awareness-based methods, or simply no contraception. Among women who do not use contraception, there may be women who are actively trying to conceive and would usually be excluded, but also those who do not mind risking a conception. Most studies also do not report asking women whether they track their fertility or menstrual cycle by counting with an app or calendar in addition to a primary contraceptive. If

women are aware of their fertility status, their answers in the fertile phase might differ spuriously due to changed behaviour (e.g. avoiding sex or using condoms, or seeking sex to conceive), heightened self-awareness for sexual thoughts and fantasies, demand characteristics, or personal theories on how their menstrual cycle affects them.

1.2.6 Directly assessing hormones may create demand characteristics

Test strips to assess ovulation via luteinising hormone surges in urine are more precise than counting methods. However, these strips are familiar to many adult women, making it easy for them to infer that a study employing these strips aims to assess effects related to ovulation and conception risk. If the participants are undergraduates at the same institution as the research team, they may accurately guess the researchers' hypotheses and consciously or unconsciously change their responses (Harris, Chabot, & Mickes, 2013). Similar worries are justified when oestrogen and progesterone are measured in saliva, blood, or urine and if women are invited back to the lab based on their menstrual cycle. In an online diary study, the study intention can be kept opaque to participants, or at least less dominant in participants' minds, especially when many other items are included. In our study, one benefit presumably was that our laboratory had not yet published research on ovulatory changes.

1.2.7 Lack of control group

Changes in oestrogen and progesterone levels around ovulation are usually hypothesised and sometimes tested as the mediating mechanism for observed

changes mid-cycle (Roney & Simmons, 2013, 2017). Unfortunately, many studies exclude women using hormonal contraceptives (HC) from taking part or from analysis, even though they can serve as a quasi-control group that experiences menstruation but not ovulation and the concurrent hormonal changes. A quasi-control group is also useful as an empirical baseline for the false discovery rate: if researchers found as many 'ovulatory' changes among HC users as among naturally cycling women, this would serve as feedback that the analysis procedure might entail false positives or invalid conclusions about the hormonal processes driving the changes. Apart from being a helpful methodological feature, including HC users allows researchers to more directly test whether, say, shifts in mate preferences or extra-pair desire do not happen among HC users. This may, simply put, be highly relevant for the many women who use HC and who might consider the absence of ovulatory cycle shifts desirable or undesirable side effects (Alvergne & Lummaa, 2010).

1.2.8 Ecological validity may be lacking

In Western societies, although female infidelity is not uncommon, with a 12-month prevalence of 2-4% and an occurrence of 20-25% per marriage (Fincham & May, 2017), few women have children with an extra-pair mate (1-2%, (Larmuseau, Matthijs, & Wenseleers, 2016)). This makes it difficult to collect the data necessary to ascertain that ovulatory shifts in extra-pair sex lead to offspring with increased fitness. Still, few instances may suffice to exert the necessary selective pressure, the low rate may be a evolutionarily recent cultural innovation (Larmuseau et al., 2016), and there has been some evidence against nonadaptive explanations of extra-pair mating in women

(Zietsch, Westberg, Santtila, & Jern, 2015). Still, most studies, lab and field, were conducted chiefly in western, educated, industrialised, rich, democratic populations (Henrich, Heine, & Norenzayan, 2010) and ours is no exception. Many studies on GGOSH have further issues with ecological validity, because women rate artificial stimuli, like morphed pictures of men, in the laboratory without consequences to their love lives. These male stimuli may highlight certain characteristics and display them in a way that exaggerates the variation from which the sampled women usually choose. Thus, effects may be overestimated and responses may not map to mate choice in the real world.

2 The present study

In the present study, we sought to replicate central findings on cycle shifts in extra- and in-pair desire, attractiveness, clothing choices and competitiveness while also improving on methodological shortcomings in the cycle research literature. By preregistering our study and main analysis plan before data collection, we reduced our own researcher degrees of freedom and thereby the risk of false positives. By using an online diary with up to 35 days reported per woman, we increased our power to detect any effects and our ability to isolate them from confounders. This design also allowed us to obtain daily reports of menstrual onset, avoiding recall error, and to do backward-counting from actual next onsets, decreasing error in the estimation of conception probability. Because diaries were filled out on participants' personal electronic devices we could assess women's reported behaviour and experiences close in both place and

time to actual behaviour. We automated the study process, decreasing our own ability to influence women's participation and responses. Because there was no cost per participant we recruited a large sample and included women regardless of contraception status, providing both a quasi-control group and making it less clear to participants what we were studying. We also assume that the automated, encrypted, minimal-contact online study made women feel more anonymous and hence comfortable to report, for instance, extra-pair desire and sex. However, using this approach implied we could not directly measure hormones, obtain photos of women, or collect ratings by their partners.

Because there is little agreement on best practices and standard operating procedures for doing this research (Blake et al., 2016; Gangestad et al., 2016; Gildersleeve et al., 2014b), we also used a variety of robustness checks to test the consequences of different decisions during data processing and statistical modelling, especially conception probability estimation, exclusion criteria and control variables.

2.1 Preregistered hypotheses

We registered the following hypotheses on the Open Science Framework on the day that data collection began. We reworded and reorganised them slightly here for space and clarity.

1. Ovulatory changes (increases during fertile window among naturally cycling women in a heterosexual relationship, but not for hormonal contraceptive users) occur in

    1. female extra-pair desire and behaviour

    2. female in-pair sexual desire

    3. having and initiating in-pair sexual intercourse (if circumstances allowed, e.g. partner was close by)

    4. subjective feelings of attractiveness

    5. choice of clothing (self-rated on the dimensions “sexy”, “figure-hugging”, “seductive”)

    6. reported male partner mate retention tactics

    7. narcissism on both dimensions of the NARC (admiration and rivalry)

2. Moderation or *shift* hypotheses: The ovulatory increase in women’s extra-pair desires and reported male mate retention behaviour is strongest (and the in-pair desire increase is weakest) for women who perceive their partners

    1. as low in sexual and physical attractiveness

    2. as low in sexual attractiveness relative to long-term partner attractiveness

    3. as less attractive compared to themselves

3. Predicted ovulatory changes are larger than, and independent of, potential ovulatory shifts in self-esteem.

In addition, we preregistered to test extraversion (4.1.), shyness (4.2.) and neuroticism (4.3.) as potential ovulatory change moderators. We called these

moderators exploratory in the preregistration to differentiate them from those already tested in the existing literature. We expected that the ovulatory increase in extra-pair desire (e.g. desire to attend social gatherings where they might meet men) may possibly be stronger for extraverted/outgoing than for introverted/shy women. Further, we expected that neuroticism may influence strength of the ovulatory increase in extra-pair desires and subjective feeling of attractiveness, though we did not specify a direction (4.4.).

3 Methods

3.1 Study description

3.1.1 Power analysis

Because we used multilevel analyses for our within-subject data, we conducted simulations to assess our study's statistical power. We simulated data under a number of different scenarios, varying among others the effect size associated with conception probability, the sample size, the number of days sampled per participant, the standard deviation of the day of the ovulation (i.e. by how much our estimated conception probability missed the correct day on average), the trait component of the outcome, and whether participants were scheduled for sampling on predicted fertile vs. non-fertile days or on random days. We did not simulate between-subjects analyses, because these should be avoided not only because of their low power (Gangestad et al., 2016) but also for reasons of validity.

3.1.2 Researcher degrees of freedom simulation

Because researcher degrees of freedom have been discussed as a source of problems in the literature, we repeated our power analysis with an effect size of zero and the following procedure simulating a hypothetical researcher engaging in the following *questionable research practices*: a) optional stopping (stop 20 or 10 participants earlier if $p < .05$), b) control for an irrelevant covariate if $p > .05$, c) try up to five correlated items as outcomes, d) start with a continuous predictor, then try broad and narrow window if $p > .05$ and combinations of these practices and determined the number of false positives.

3.1.3 Preregistration

We preregistered our study's hypotheses and methods on March 19, 2014 and added a planned amendment to our exclusion criteria and fertility estimation method to the preregistration on May 10, 2014, when data collection was already underway (Schilling, Straus, Arslan, Gerlach, & Penke, 2014). Participants enrolled from March 19, 2014 to July 2, 2015. The last diary entry was made on December 3, 2015. The preregistration had a second part, which pertained to hypotheses related to oestrogen dosage effects in hormonal contraceptives and which we plan to discuss in a separate manuscript.

In our initial preregistration, we specified that we would use backward counting from the observed next menstrual onset to estimate a narrow fertile window (reverse cycle days 15-19 vs. 2-11). After the publication of Wood et al. (2014), we decided to also test

a broad window (reverse cycle days 14-22) in order to compare results using the two approaches. Moreover, we preregistered that we would descriptively show results based on continuous curves centred on the estimated day of ovulation. We preregistered the personality and daily diary items we would use. For sample size calculations, we did not preregister a fixed sample size, as this is hard to control in online studies and power analyses based on a biased literature are of limited use. Instead, we preregistered a complex procedure under which we tried to ensure that we would obtain an adequate sample size even if recruiting proceeded slowly and that students could finish projects based on this study in a reasonable timeframe. We stopped recruiting when we were unable to find further participants and we honoured our promise not to stop data collection depending on results.

3.1.4 Participants

We recruited women via university mailing lists in Germany, newspaper articles about our group's work (without references to ovulation-related work), our online study site psytests.de, word-of-mouth, and among local students in exchange for course credit at our university. Only participants who self-reported their sex as female and reported currently being in a heterosexual relationship were allowed to participate. Out of the 1,720 participants who signed up for the study, 259 were ineligible to participate according to these criteria, 253 did not complete the demographics and personality survey preceding the diary, 54 completed no diary entries, 41 were sterilised, infertile or pregnant, and fertility was never estimable for 70 due to few or patchy diary entries. Out of the remaining participants, 60% (n = 631) were using some form of hormonal

contraceptive and 40% (*n* = 428) were naturally cycling. Specifically, 5% (*n* = 53) used a fertility-awareness-based-method, 28% (*n* = 291) used only barrier methods, mostly condoms, and 6% (*n* = 67) reported no contraception. We preregistered several exclusion criteria that we deemed useful to exclude women with potentially anovulatory cycles. Applying the strictest criteria proved to be over-exclusive, as only 13% of the naturally cycling sample would have been retained. Hence, we differentiated our exclusion criteria into four strictness levels and examined the effect of applying these levels in robustness checks. The participant flow and exclusion criteria are shown in Figure 1.

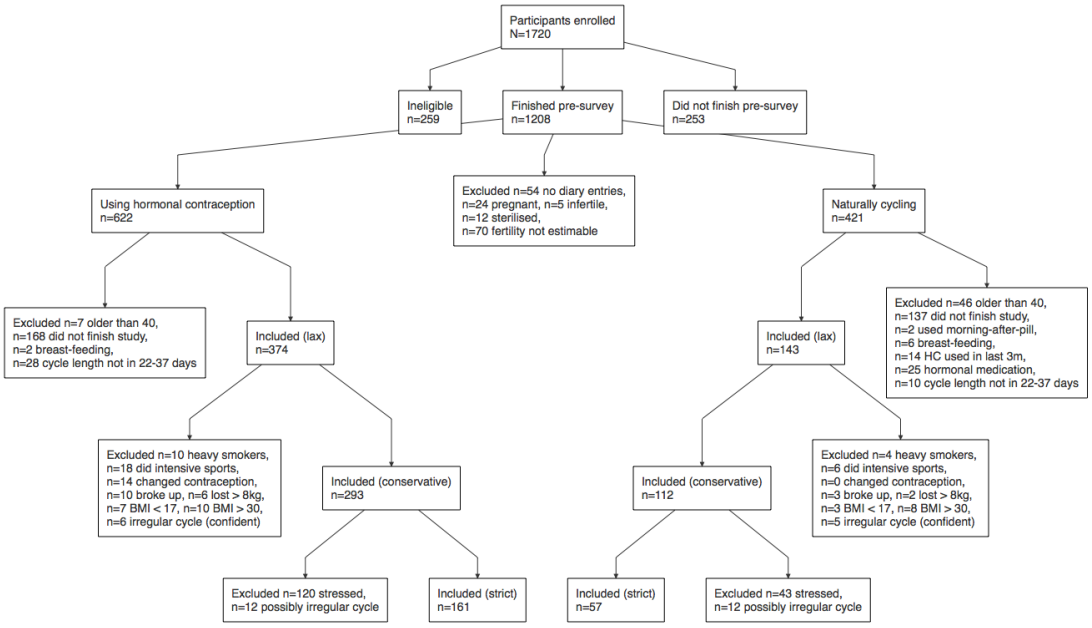


Figure 1. Participant flow. The figure depicts the various exclusion criteria and the number of participants affected by each (if not already excluded for a preceding reason).

The 1,043 eligible participants were on average 25.5 years old (*SD* = 6.3, 18-53) and had been in a relationship for 3.8 years (*SD* = 4.3). Most (71%) were students, 24%

were working, 3% were not working or described themselves as homemakers, and 3% were in secondary or vocational school. A majority reported their religious denomination as Christian (56%) and 42% described themselves as nonreligious. Twelve percent were married and a further 4% engaged to be married. Four percent of the sample reported not yet having had sex with their partner. Most (88%) had no children. The largest group co-habited with their partner (41%), but a sizeable fraction had a long-distance relationship (31%), with the remainder living in the same city. Of those who did not live with their partner, 34% lived in a flatshare and 25% lived alone. We present more detailed data on the distance between partners, how often they saw each other and spent the night in the online supplement. Geographically, only our university town seemed visibly overrepresented. Hormonal contraceptive users differed from naturally cycling women in a number of ways (see Table 1 for continuous variables and online supplement for all others). Most importantly, they were almost 5 years younger on average, and consequently tended to be unmarried and not to co-habit, to be in relationships for a shorter time (approximately 2 years), to have had 3.5 fewer lifetime sexual partners, to be students and have lower income. However, when simultaneously predicting hormonal contraception status from 28 demographic and personality predictors in a probit regression, only lower age, low openness, high conscientiousness, and being unmarried were significantly predictive at $p < .05/28$. For the sample used in our preregistered analyses, the only differences remaining significant in the regression were that women on the pill were approximately 3 years younger and lower in

openness. Hormonal contraceptive users also had shorter and more regular cycles, which might be consequences rather than causes.

Table 1. Descriptive statistics by hormonal contraceptive use.

	Mean (Standard deviation)			
Variable	**HC user**	**Cycling**	**Hedges' g**	***p***
Age	23.6 (4.4)	28.4 (7.6)	1.10	< .001
Religiosity	2.0 (1.1)	2.0 (1.2)	0.01	.891
Age at first time (years)	16.9 (2.3)	16.9 (2.4)	-0.01	.886
Age at menarche (years)	13.0 (1.3)	13.0 (1.5)	-0.06	.557
Relationship duration (years)	2.9 (3.0)	5.0 (5.5)	0.70	< .001
Cycle length (days)	27.9 (2.9)	29.1 (3.6)	0.41	< .001
Life no. sexual partners	5.7 (7.2)	9.3 (14.9)	0.50	< .001
BFI Extraversion	3.5 (0.8)	3.5 (0.8)	0.03	.638
BFI Agreeableness	3.6 (0.6)	3.6 (0.6)	0.00	.964
BFI Neuroticism	3.1 (0.7)	3.0 (0.8)	-0.14	.037
BFI Conscientiousness	3.6 (0.7)	3.5 (0.7)	-0.15	.024
BFI Openness	3.6 (0.6)	3.8 (0.6)	0.31	< .001
Relationship satisfaction	4.2 (0.7)	4.0 (0.8)	-0.20	.003

Notes. Constructs in bold remained significant after multivariate adjustment in a probit regression. BFI = Big Five Inventory. HC = hormonal contraceptive.

3.1.5 Procedure and implementation

3.1.5.1 Procedure

3.1.5.1.1 Intake form and consent

Participants filled out web-based questionnaires on their personal electronic devices (27% used a mobile device). They were informed that the study's purpose was to

examine the relationships between everyday life, relationship events, psychological well-being, and sexual behaviour. They were told that each diary day they filled out would add one more lot in a lottery for four Amazon.com coupons worth 20€ each and that they would receive extensive feedback on their personality and the longitudinal co-development of their mood, self-perceptions, and clothing choices over week days. Students of our university could earn course credit instead. They were informed that, although the study required their email address to send diary invitations, data would be stored separately and anonymously and that the feedback would also be generated anonymously and automatically. Research that only entails self-reports does usually not require IRB approval under German regulations.

3.1.5.2 Demographic and personality survey

After obtaining consent, we asked participants for their sex, age, and relationship status. Only self-identified females in a heterosexual relationship could proceed. Next, the women reported various demographics, details about their relationship, their menstrual cycle and contraception status and completed several measures of personality, relationship satisfaction and jealousy (see Table 2).

3.1.5.2.1 Diary

On the next day and until at least 30 entries were obtained over a period of at least 40 days, women were invited to fill out the diary via email and, if possible, text message at 5 pm German time. They could fill out the diary until 7 hours after the invitation was sent. Participants completed the diary in a median time of 6.5 minutes. In each diary entry, they responded to 58 items about their relationship, interactions with their partner,

clothing style, self-esteem, narcissism, sexual desire and behaviour, and menstrual cycle (see below). They were asked to refer to the period since their last entry or 30 hours ago, whichever happened sooner. They could also give free-text responses to provide context for their entry.

3.1.5.2.2 Follow-up survey

After completing the diary (usually immediately after the last day), women were invited to a follow-up survey. In this survey, we asked several questions which we expected to relate to the validity of the results, namely what they thought the purpose of the study was, whether they were ill, took medication, lost weight, smoked, broke up with their partner, started a new relationship, switched contraception methods, or felt extraordinarily stressed. They then received their feedback. If they had not menstruated during the last 14 days of the diary, we sent them reminders every other day inviting them to tell us about their next menstrual onset, continuing until they did.

3.1.5.3 Implementation

The study was implemented using the online open-source survey framework formr.org (Arslan & Tata, 2016). The software permitted us to automate all repetitive aspects of the study, such as administering surveys, sending email and text message invitations and to generate graphical feedback for participants. The study administrators communicated with participants through an email account and could send manual reminders and administer service requests in case of problems without seeing the participants' data.

3.1.6 Measures

We documented all items for all surveys in the online supplement. To assess reliability for cross-sectional measures we computed Cronbach's alpha. For within-subject measures, we computed the generalizability of within-subject change aggregated across items (Shrout & Lane, 2012) using the *psych* package (Revelle, 2017). We documented the main outcome measures for the diary and their reliabilities in Table 2. We used measures from previous studies where possible, but previous studies often could not or did not test the relevant generalizability metric for a within-subject process, namely whether the scale measured within-subject change reliably. Unfortunately, this did not appear to be the case for the mate-retention-related measures, and generalizabilities for the other outcomes were lower than optimal. The cross-sectional measures of personality, i.e. the Big Five Inventory (Lang, Lüdtke, & Asendorpf, 2001) and shyness (Asendorpf & Wilpers, 1998), had Cronbach's *α*s ranging from .81 to .88. Agreeableness, which we did not use in this study, was an exception with α = .73. Confidence intervals (95%) for these *α*s had a width of 0.02-0.04. The reported physical attractiveness of the partner was based on two items (taken from (Haselton & Gangestad, 2006)) asking about his physical attractiveness and his sexiness (α = .80). The reported short-term attractiveness of the partner included the physical attractiveness scale, plus an item about his attractiveness for an affair or one-night stand and an item asking about sexual satisfaction with this partner (α = .62). To compute the partner's attractiveness relative to oneself (Haselton & Gangestad, 2006) we first computed a five-item mate value scale (Landolt, Lalumière, & Quinsey, 1995) for the partner and the participant. We omitted two items in both scales because they

used tortuous sentences and counterfactuals and exhibited low to negative scale loadings. Own mate value (α = .84) correlated .25 with partner mate value (α = .78). We then tested whether the four-point Likert item "Who does better with the opposite sex? You or your partner?" favoured the partner most when his mate value exceeded hers. This was the case. Thus, we standardised and summed the mate value difference and the latter item (α = .74). The relative measure was uncorrelated with the various absolute measures (|r| < .05). Further details on scale construction and reliabilities are available online. Confidence intervals (95%) for *α*s of the attractiveness-related scales had widths from .04-.07.

Table 2. Outcome measures in the diary.

Construct	Scale Origin	Items	Rcn	Example item
Female Jealousy		3	.00	"I have asked my partner with whom he spent his day."
Relationship satisfaction		1	.85	"How satisfied were you with your relationship?"
"Sexy" clothing	Schwarz & Hassebrauck, 2008	3/8	.60	"Would you describe your chosen clothes today as sexy?"
Extra-pair desire	Haselton & Gangestad, 2006	12	.60	"I had sexual fantasies about men other than my partner."
Partner mate retention	Haselton & Gangestad, 2006	4	.00	„My partner asked my with whom I spent my day."
Female mate retention	Haselton & Gangestad, 2006	6	.17	"I told my partner I love him."

Narcissistic admiration and rivalry	NARQ-K (Back et al., 2013)	3+3	.57/.55	"I felt worthy of being seen as a great personality."
Self-esteem	RSES Rosenberg, 1965	1	.86	"I was satisfied with myself overall."
Self-perceived desirability		1	.85	"I felt sexually desirable."
In-pair desire		3	.75	"I found my partner particularly sexually attractive."

Notes. Rcn = Reliability of change or generalizability of within-person variations. For clothing choices, three of eight items asked about "sexy" clothing choices.

3.2 Analysis

3.2.1 Menstrual onset computation and fertile window inference

On each diary day, women reported whether they had had their period on that day or in the preceding 6 days. As this meant that women could report the same menstrual onset multiple times and hence incorrectly recall a menstrual onset a few days later, we always used the report closest to the reported onset. Women also reported a last menstrual onset in the survey preceding the diary and a next menstrual onset in a follow-up survey after the diary. We used these dates to generate time series for each participant. We then counted forward and backward from each menstrual onset to the next or respectively last menstrual onset. If the next menstrual onset was not available, because women did not complete the follow-up survey, we could infer it from the reported average cycle length, but only did so for our robustness checks. We then inferred a narrow and a broad fertile window. For our robustness analyses, we

additionally computed a continuous estimate of the probability of being in the fertile window according to the method advocated by Gangestad et al. (2016), who based their estimates on (Stirnemann, Samson, Bernard, & Thalabard, 2013), among other data. This method accounts for the fact that the luteal phase length is less variable than the follicular phase. Further details can be found in the online supplement. This procedure resulted in seven different predictors which allowed us to include a varying number of diary days, see Table 3.

Table 3: The different conception probability estimates that were used as predictors.

Description	fertile window	n (days)	% of days	n (women)
all days		28,493	100	1043
narrow window, backward counted	15-19	9501	33.35	794
broad window, backward counted	14-22	11,497	40.35	796
narrow window, forward counted	11-15	12,171	42.72	973
broad window, forward counted	8-16	15,880	55.73	997
continuous, backward counted	n/a	17,614	61.82	817
continuous, backward counted from reported cycle length	n/a	26,580	93.29	1043

Notes. To make effect sizes across predictors comparable, we dummy-coded windowed predictors as being 0.053 on non-fertile days and 0.44 (broad)/0.51 (narrow) on fertile days. These were the averaged probabilities for those days from the continuous estimate, which varied from 0.01 to 0.58. Days were counted from the menstrual onset, starting at 1. The non-fertile window was defined as days 4-12 (backward-counted) or respectively days 18-26 (forward-counted).

3.2.2 Statistical approach

To test our hypotheses we fitted multilevel models in *lme4* (Bates, Mächler, Bolker, & Walker, 2014) with a random intercept per person, interacting our fertility estimate with a dummy for hormonal contraceptive use. Defining the model in this way allowed us to both test whether any ovulatory change among naturally cycling women was different from zero, as well as whether it was different from any changes occurring among hormonal contraception users. For Likert-scaled outcomes we fitted linear multilevel models and for categorical outcomes we fitted generalized linear multilevel models with a binomial family using a probit link. In Wilkinson notation (Bates et al., 2014, p. 4; Wilkinson & Rogers, 1973), the model equation can be formalised as outcome ~ fertile_window * hormonal_contraceptive_user + (1 | person)

Here, *fertile_window* refers either to the backward-counted narrow or broad fertile window in the preregistered analyses. To test H3.1 we also refitted models with self-esteem as a covariate. Because we did not preregister it, we did not fit random slopes for the fertile window effect. We instead examine the effect of doing so in our robustness checks (Bates, Kliegl, Vasishth, & Baayen, 2015).

3.2.3 Robustness checks

To test our results for robustness, we used a variety of approaches. First, we built a baseline model that deviated from our preregistered procedure but implemented the best practices published after we preregistered (Blake et al., 2016; Gangestad et al., 2016).

In Wilkinson notation, the model can be formalised as

outcome ~ (*fertile_window_probability* + premenstrual_phase + menstruation) * hormonal_contraceptive_user + average_fertile_window_probability+ (1 | person)

Here, the probability of being in the fertile window was continuously estimated from backward counting from the next menstrual onset, according to Gangestad et al. (2016). In cases where the next menstrual onset was not observed, we fell back to the next menstrual onset inferred from the average cycle length that women reported in the screening survey (see Table 3). Because using a continuous predictor means that days on which women were menstruating or in the premenstrual phase were also included, we included dummy variables for the reported menstruation and the inferred premenstrual phase (the six days before the menstrual onset). We also adjusted for average probability of being in the fertile window per woman as an additional predictor, to ensure within-person estimates (Bafumi & Gelman, 2006). We let our fertility and menstruation predictors interact with hormonal contraception status.

In this baseline model, we included all usable data (from 1,043 women, 421 naturally cycling) instead of excluding many women based on our preregistered criteria.

We tested robustness by fitting numerous variations on the baseline model described above. We then examined the effect size and standard error of the fertile window predictor across many models, which we outline in the following.

In model *M_r1*, we allowed a varying slope per participant for the fertile window and the two menstruation dummy variables, a “maximal” specification that is somewhat

controversial because of the potential for overparameterisation (Barr, Levy, Scheepers, & Tily, 2013; Bates et al., 2015). We tested four levels of stringency for exclusion ("all", "lax", "conservative", "strict", see Fig. 1) in models *M_e1-4* and *M_m5*. We also tried to implement a post-hoc criterion (*M_e5*) for data reliability, under which we excluded 1251 diary days (4% of all) where participants a) gave the same answer to all Likert items (n=23) or b) accessed the diary later or earlier than intended due to technical problems (n=896) or c) took more than 24 hours (n=376) or less than a minute (n=30) to finish filling out the diary. We took these steps to reduce the number of careless responses and to remove days on which the assigned cycle day might be off. We also tested (*M_e6*) whether the effect of excluding women who were trying to get pregnant, an exclusion criterion we had not preregistered.

In models *M_p1* to *M_p11*, we tested different estimates of the fertile window as our predictor to address the concerns described in section 2.2.2. We compared all combinations of a narrow window, broad window, continuous estimates, and backward- and forward-counting. To address section 2.2.3 and 2.2.4 empirically, we then tested whether effects could be shown using only a single day per participant, two days (at low and high fertility) or four days (two each).

To transparently show how much modelling decisions that might be considered researcher degrees of freedom (section 2.2.1) matter, we fitted models *M_c1* to *M_c5*. In these, we added adjustments in one model each for *M_c1* self-esteem, *M_c4* week day and week number, and *M_c5* the time when the diary was started and how long it took to fill out, or we omitted adjustments for *M_c2* average fertile window probability, or

M_c3 both average fertile window probability and menstruation. This allowed us to see the effect these adjustments had on the estimated fertility effect. In *M_c6* to *M_c7*, we tested two different temporal autocorrelation models as opposed to the unstructured random effect correlations in our main model. In M_c9, we tested whether measurement reactivity might confound our results, by adjusting for splines for the number of days since the diary beginning (a variable for days filled out and one including missing days), by hormonal contraceptive use.

We then tested various moderators to prod different methodological issues. To partially address the issues pointed out in section 2.2.5, in *M_m1* we compared four groups of contraceptive methods (hormonal, awareness-based, barrier-based, none). For women who combined multiple methods, the order of the list above determined precedence. To test generalizability, we tested moderation by participant age (in groups 18-20, 20-25, 25-30, 30-35, 35-45, 45 and older, *M_m2*), and whether the weekday (*M_m3*) or the weekend (*M_m4*) moderated effects (Roney & Simmons, 2013). Because the validity of fertility estimates from counting methods depends on accurate reporting and regular cycles, we tested for moderation by cycle length (*M_m6*), by self-reported certainty about menstruation parameters (*M_m7*), and by self-reported cycle regularity (*M_m8*). To further test generalizability, we also tested for moderation by cohabitation (*M_m9*) and by marital status (*M_m10*).

We also ran Bayesian regression models using Stan (Bürkner, in press; Carpenter et al., 2015) to be able to appropriately model the positively skewed distribution of the Likert items for extra-pair desire (i.e. many respondents indicated minimal extra-pair

desire) in an ordinal regression using a cumulative outcome distribution and random effects for items and participants. In the Stan models, we also tested for heterogeneity of effect sizes across participants and items. In additional Stan models, we fitted a thin-plate regression spline (S. N. Wood, 2003) over backward-counted cycle days to examine whether the continuous probability of being in the fertile window would be a good fit to the shape of the estimated effect. In exploratory analyses, we also fitted one Stan model per item and graphically summarised the posterior densities for the conception probability estimates. Because of computational limitations, we fitted models separately instead of pooling information across items and scales.

In our robustness checks, a null hypothesis testing approach would have been inappropriate, given the wide-ranging exploration and varying questions asked across outcomes and models. Instead, we focused on visualisations and the fertility effect's point estimate and confidence interval. We inspected effects to look for evidence that an effect was not robust (i.e. shifts in estimates that might not be explainable by sampling error). We summarise what we consider the main patterns, but made the detailed results available online (see below).

3.3 Data, code, results, and materials availability

We released all code, both for implementation and analysis, materials, and full statistical results pertaining to this study openly in the online supplement (https://rubenarslan.github.io/ovulatory_shifts/). We partially anonymised the data and uploaded them to the Open Science Framework for safekeeping. However, because sexual diary

data are hard to completely de-identify and extremely sensitive, we did not request consent from participants to share their data openly and cannot share these data publicly. Therefore, we can only share the partially anonymised data with anyone who has a valid reason and agrees not to attempt to re-identify the data. We have also generated a synthetic dataset using synthpop (Nowok, Raab, & Dibben, 2016). This dataset attempts to replicate many of the central features of our data, such as means and bivariate associations, but is anonymous. Others can use this to test and build models using realistic fake data, which we can then easily test on the real data.

4 Results

4.1 Power analysis and researcher degree of freedom simulation

We documented our power analyses and researcher degrees of freedom simulations and results in more detail online. They showed that under reasonable assumptions, power was a function of the number of usable days multiplied by the sample size.

To detect a regression coefficient of the fertile window of .2 with an alpha level of .01 in a sample of 150 naturally cycling women measured over 30 days, we had a power of .84 using a windowed predictor, because using windows meant not being able to use many of the measured days. Using a continuous predictor increased power to .99. In a sample of 500 women measured over 30 days, power approached 1. Power to detect an effect half/a quarter this size was still .97/.36 using a continuous predictor.

4.2 Preregistered analyses

To adjust for multiple comparisons, we set the significance level to .01 (see below). After applying our “lax” exclusion criteria (see robustness checks for further tests of stringency), we could use data from 143 naturally cycling women and 374 hormonal contraceptive users. Using the narrow (broad) fertile window predictor, we could use 6,378 (7,740) diary days, or 12 (15) days per woman (see Table 3).

All outcomes are summarised in Table 4. For three outcomes, effects of the fertile window were significantly positive for naturally cycling women but absent for hormonal contraceptive (HC) users, a pattern we will refer to as *fertile window increases* in the following. When the interaction between HC use and the fertile window is of the same size as the fertile window effect, but negative, it indicates an absence of the change among HC users.

We found small fertile-window increases in extra-pair desire and behaviour. Effects were significantly positive for all extra-pair subscales except the compliments subscale. We examined this pattern in more detail in the robustness analyses. Actual instances of intimate contact or sex with another person were very rarely reported (48 women reported extra-pair sex on 127 days, 112 women reported extra-pair intimate contact on 383 days), so that the log-odds-ratios seem large, but estimates were not significant (ps > 0.17).

We also found small fertile window increases in in-pair desire, similar in size to the increase in extra-pair desire. On average, women did not have significantly more sex

during the fertile window, but there were two consistent but only marginally significant moderators of the ovulatory increase in having sexual intercourse, namely cohabitation and average number of nights spent with the partner. Cohabitation moderated the changes, so that we observed no ovulatory increases among women in long-distance relationships ($p = .020$). Women who spent more nights per week with their partner also showed stronger ovulatory increases ($p = .048$). The increases were not stronger on the specific nights that the couple spent together ($p = .58$). Women did not initiate sex significantly more often.

We also found small fertile window increases in self-perceived desirability, but not on wearing “sexy clothes”. The predicted effects were not significant for initiating sex, male mate retention, narcissistic admiration, and narcissistic rivalry (all *p*s > 0.21).

The changes in self-perceived desirability, in- and extra-pair desire were also clearly apparent when plotting a smoothed spline over reverse-counted cycle days (Figure 2). The pattern of results held independently of whether we used a narrow or broad fertile window as the predictor. As predicted, there were no significant effects on self-esteem and adjusting for self-esteem did not weaken any other tested associations.

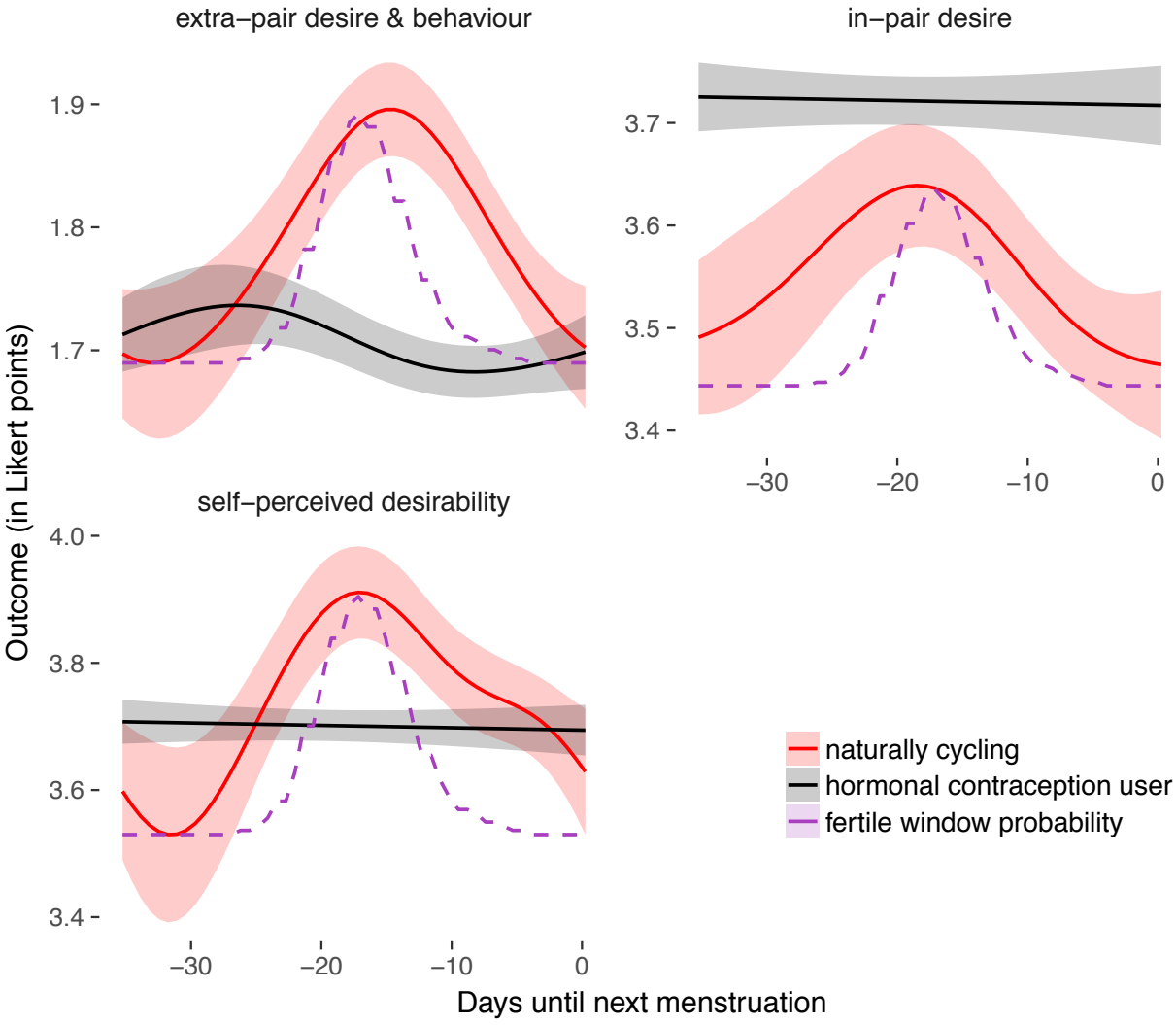


Figure 2. Smooth thin-plate splines (S. N. Wood, 2003) fitted over days until next menstruation with three central outcomes. The dashed line shows the estimated probability of being in the fertile window for each day. The shaded areas reflect 95% confidence bounds pooling days over participants for simplicity. To account for the cyclical nature of the data, we spliced in duplicates of the time series at both ends before estimating the splines and then dropped them afterwards.

None of the three main predicted moderators, i.e. the partner's short-term, sexual, and relative attractiveness, significantly exhibited the predicted pattern for any outcome (*p*s > 0.07), and some patterns went descriptively in the opposite direction of the prediction. Also, none of the personality variables moderated changes in extra-pair desire and behaviour (p*s* > .32). A test of whether neuroticism moderated shifts in self-perceived desirability was significant (p = .002), but inspection of marginal effect plots showed this to be driven by significant increases in desirability among highly neurotic hormonal contraceptive users, an unpredicted and likely spurious result.

Because we had not preregistered a procedure to correct for multiple comparisons due to multiple outcomes and believed Bonferroni to be too conservative, as many outcomes were highly correlated, we tested whether we would have ever rejected the null hypothesis of no effect in our HC control group with the significance threshold of .01. Although this would have been the case for one outcome, follow-up analyses showed that this result would not have survived our robustness analyses, so we concluded that our chosen threshold was appropriate. The pattern of significant results here would not have been different using the uncorrected threshold of .05 or when using a Benjamini-Hochberg (Benjamini & Hochberg, 1995) correction (see online).

Table 4. Preregistered associations, using the narrow fertile window

Outcome	Intercept	fertile	HC user	HC user x fertile
Extra-pair desire and behaviour				
extra-pair (EP)	1.75±0.05	**0.27±0.06**	-0.05±0.06	**-0.30±0.07**

desire & behavior		$p < .001$	$p = .373$	$p < .001$
- EP compliments	2.37±0.08	0.25±0.11	-0.11±0.10	**-0.37±0.13**
		$p = .023$	$p = .267$	$p = .005$
- EP flirting	1.36±0.04	**0.15±0.06**	-0.09±0.05	**-0.22±0.07**
		$p = .006$	$p = .078$	$p < .001$
- EP going out	1.99±0.09	0.24±0.15	0.24±0.10	-0.31±0.18
		$p = .113$	$p = .019$	$p = .088$
- EP sexual fantasies	1.50±0.06	**0.49±0.09**	-0.19±0.08	**-0.43±0.11**
		$p < .001$	$p = .012$	$p < .001$
- EP desire	1.65±0.05	**0.34±0.06**	-0.13±0.06	**-0.31±0.07**
		$p < .001$	$p = .047$	$p < .001$
extra-pair intimacy[pb]	-4.47±0.30	0.89±0.42	-0.22±0.37	-0.57±0.72
		$p = .033$	$p = .554$	$p = .431$
extra-pair sex[pb]	-4.60±0.39	0.60±0.56	-0.44±0.57	0.17±1.08
		$p = .282$	$p = .444$	$p = .873$
In-pair desire and behaviour				
in-pair desire	3.48±0.08	**0.31±0.12**	0.24±0.09	**-0.39±0.14**
		$p = .010$	$p = .010$	$p = .008$
sexual intercourse[pb]	-0.98±0.07	0.12±0.17	0.17±0.08	-0.26±0.20
		$p = .483$	$p = .026$	$p = .203$
sex initiated by partner vs. woman[pb]	0.26±0.09	-0.14±0.31	0.12±0.11	0.11±0.37
		$p = .642$	$p = .276$	$p = .775$
partner mate retention	2.86±0.07	0.05±0.09	0.00±0.08	-0.12±0.11
		$p = .569$	$p = .954$	$p = .255$
Self-perceived desirability and clothing choices				
self-perceived desirability	3.72±0.08	**0.37±0.13**	-0.07±0.09	-0.38±0.15
		$p = .004$	$p = .477$	$p = .012$
sexy clothing	3.16±0.07	-0.14±0.10	0.02±0.08	0.09±0.12
		$p = .169$	$p = .831$	$p = .492$
Narcissism				
narcissistic admiration	2.69±0.10	-0.05±0.08	-0.14±0.11	-0.09±0.09
		$p = .551$	$p = .214$	$p = .335$
narcissistic rivalry	1.29±0.04	-0.03±0.05	0.05±0.05	-0.02±0.06
		$p = .535$	$p = .322$	$p = .747$

Notes. Coefficients significant at $p < .01$ (before rounding) are bold. Associations with outcomes marked [pb] were estimated in a probit regression. The number after the ± is a standard

error. Scales starting with EP are subscales. The sex initiation item asked whether it was rather the partner or rather the participant who initiated sex, in a forced-choice question. Positive effects reflect that it was rather the partner.

4.3 Robustness checks

Our robustness check results are documented fully in the online supplement, here we verbally and visually summarise the most important patterns. We were able to include 421 NC women and 622 HC users. We used a continuous measure of the probability of being in the fertile window, estimated from backward-counting from the actual next menstrual onset, falling back to the next menstrual onset inferred from the average cycle length when necessary. This way, we were able to include 25,948 diary days, i.e. on average 25 days per woman and more than 3 times as many days as in the preregistered analyses.

We repeated all preregistered tests using this bigger dataset and the adjusted model. Unless otherwise mentioned, results were robust to including more data and to the various checks listed in the method section. Specifically, estimates of fertile window increases in extra-pair desire and behaviour, in-pair desire, and self-perceived desirability were robust, but standard errors shrunk by about half. Further, none of the predicted moderation patterns turned significant when adding more women, and using slightly different items for the partner attractiveness moderator variables did not change the pattern. However, when modelled, random slopes for the fertile window predictor were substantial, larger than for the menstruation predictors and as large as the residual variation and the variation explained by the random intercept. No fertile window

increases emerged for any other outcomes, including further outcomes for which had not predicted increases.

We found that the stringency of our exclusion criteria, designed to exclude women with potentially anovulatory cycles, did not moderate the effect sizes in the expected way, i.e. that effects became stronger with more stringent criteria. When testing for moderation by exclusion criteria in *M_m5*, the pattern validated our post-hoc decision to keep only the truly necessary constraints. When applying stricter exclusion criteria, some effects weakened or confidence intervals overlapped zero, but this seemed to reflect the heavily decreased sample size (see Figure 1). Applying our post-hoc criterion (*M_e5*) to exclude potentially unreliable data also had no noteworthy effect. Excluding women who were trying to get pregnant diminished the effect on in-pair desire, but did not eliminate it.

When we used a continuous fertile window predictor, we also adjusted for premenstrual and menstrual days. We found that including adjustments for menstruation and pre-menstruation (*M_c3*) reduced effect sizes for the fertile window predictor. We could not always adjust for menstruation when using a narrow window predictor because of model convergence problems. After taking this into account, we found no systematic pattern in which certain predictors (narrow or broad window, forward or backward counted) had larger effect sizes than others across outcomes (see Figure 3). However, continuous curves over backward-counted days (Figure 2) matched

the predicted pattern more closely than curves over forward-counted days (online).

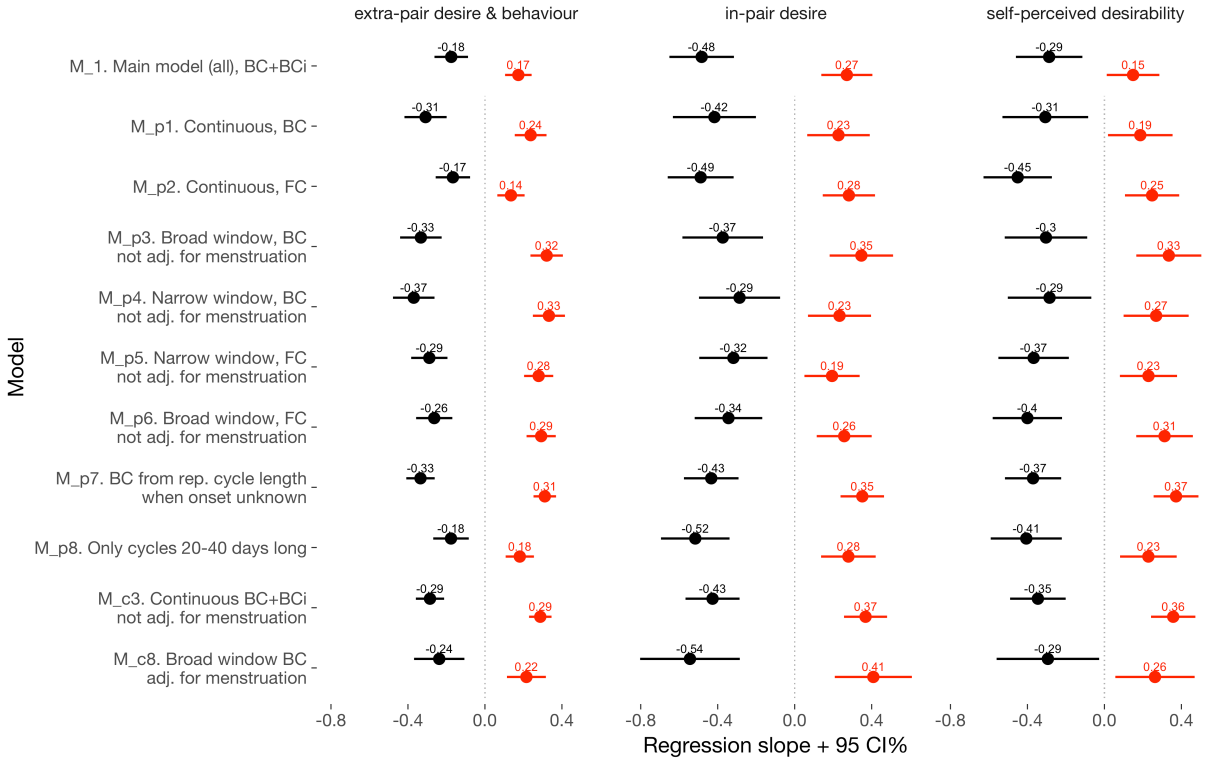


Figure 3. Coefficient plot showing a consistent effect of the fertility predictor among naturally cycling women (red) but not HC users (black) across several predictor and model specifications (explained in further detail in the text).

In models *M_p7 to M_p9,* we found that none of the associations found to be significant in the pre-registered analyses would have been discovered had we used between-subject analyses or a high-low fertility within-subject design with only two days.

There was a complex pattern of results when separately examining contraception methods. The ovulatory increase in extra-pair desire tended to be larger for fertility-aware women (5% of the sample) and this was not merely because they had more regular cycles. Still, women using barrier methods or no contraception also showed

ovulatory shifts. The shifts in in-pair desire and self-perceived desirability, on the other hand, appeared weaker or absent in fertility-aware women but stronger in women using no contraception (6% of the sample). Because women using methods other than hormonal contraceptives and barrier methods made up only a small minority of the sample, we could not rule out sampling variation as an explanation.

Inspecting time series of within-subject change by item (Figure 4) for the three outcomes that were significant in the preregistered analysis, namely extra- and in-pair sexual desire and self-perceived desirability, showed that naturally cycling women tended to exhibit peaks around the estimated day of ovulation, while hormonal contraceptive users exhibited no clear peaks or minor peaks around menstruation.

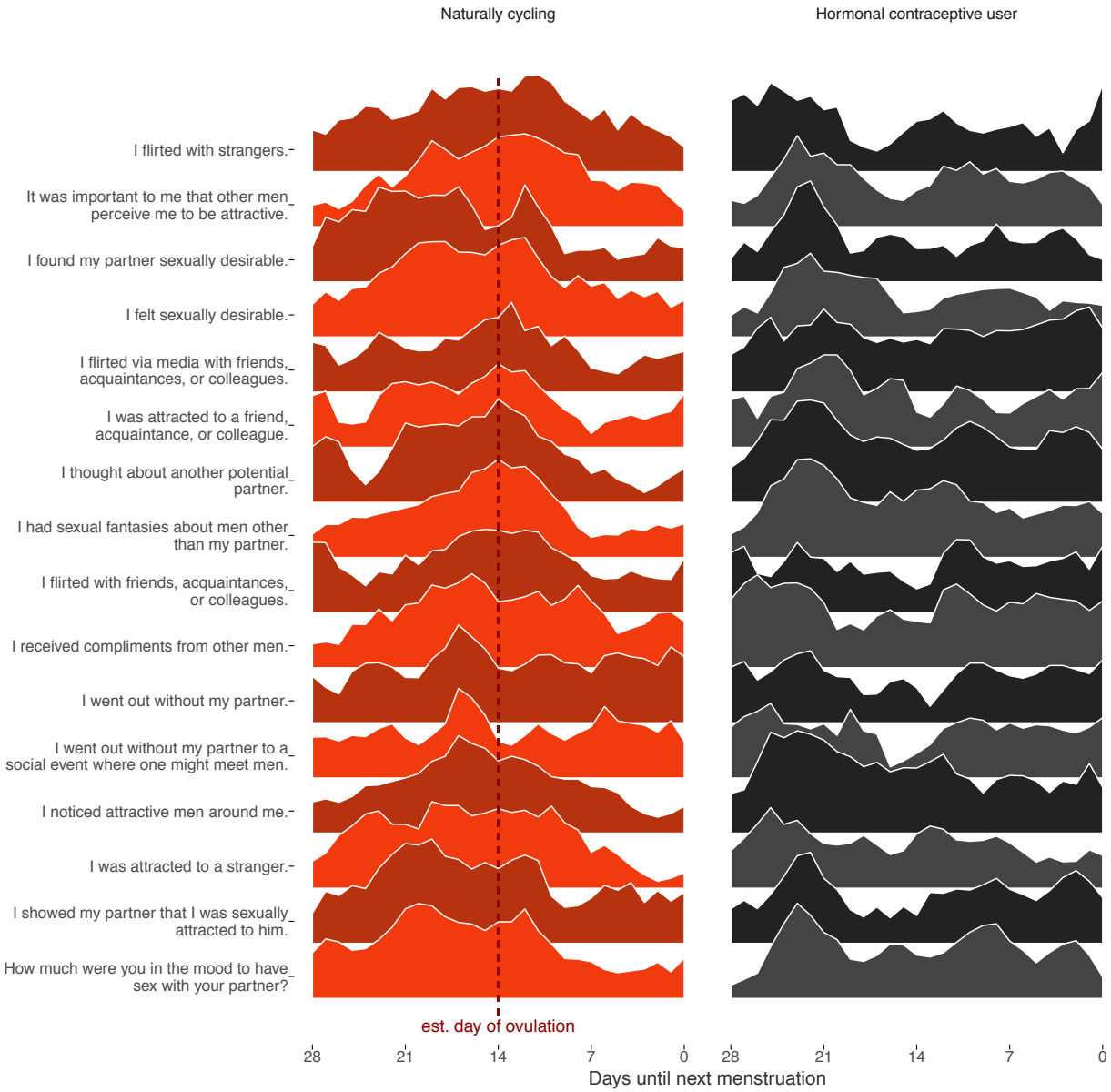


Figure 4. Item-by-item plot of within-subject change. The trails in this plot represent within-subject change as a percentage of the maximal peak. Plots are smoothed with a moving average over three days. Items are ordered top to bottom by how late in the cycle the highest peak occurs for naturally cycling women.

Across outcomes, effects tended to be largest for women with cycle lengths between 25 and 30 days, and for women who were more certain about their menstruation parameters, but not for women whose cycles were more regular.

We tested whether the effects of in- and extra-pair desire were different in size and independent of each other, to test whether they were potentially both driven by a third variable, such as increased target-unspecific sex drive. The two categories of desire negatively correlated within each woman, so that adjusting either desire outcome for the other did not diminish the estimated fertile window increase. However, we also conducted simple forward simulations of the realistic scenario that unobserved properties of the object of desire decide whether desire is expressed as in- or extra-pair. These simulations showed that we cannot resolve the question of whether the effects were entirely explained by target-unspecific desire without directly measuring it. Comparing unstandardized effect sizes showed that the fertile window increases in extra-pair (b= 0.26 95% CI [0.17;0.35]) and in-pair desire (0.26 [0.10;0.42]) were comparable in size. Examining item-level effect sizes showed larger heterogeneity across items than across objects of desire (see online supplement).

5 Discussion

In the present large diary study, we aimed to replicate reports of ovulatory changes in extra- and in-pair sexual desire and behaviour, as well as related outcomes, and test several methodological concerns. We could replicate only some of the previously reported ovulatory changes, namely those in the three main outcome categories of

extra-pair sexual desire and behaviour, in-pair sexual desire and behaviour, and self-perceived sexual desirability. In Figure 2 we show the probability of being in the fertile window closely matches the observed changes across the cycle for these three outcomes.

5.1 Main effects of the fertile window

5.1.1 Extra-pair desire and behaviour

We found robust support for a fertile window increase in extra-pair desire and behaviour. This scale was a fairly heterogeneous average of items measuring increased attraction to, fantasizing about, flirting with, receiving compliments from, and going out to meet with men other than the primary partner. In separate analyses, we also examined whether women were more likely to be intimate or have sex with other men during the fertile window. While descriptively supporting the predicted ovulatory shifts, these events were rare and effects were not significant. We also examined effects on the subscale level. Fertile window increases in sexual fantasies were descriptively strongest, but the aggregation of subscales seemed justifiable.

5.1.2 In-pair desire and behaviour

We found robust support for fertile window increases in in-pair desire. Although in-pair desire predicted intercourse with the partner, ovulatory increases in sexual intercourse were not significant in our preregistered analyses. Potentially, we simply had too little power to detect mean shifts in this dichotomous behaviour: Women

reported sex on 21% of days and 67 women who filled out the diary on more than 25 days never reported sex with their partner at all. With added data, we observed increases in some of our robustness tests, but only in comparison to the HC group (which decreased non-significantly). Further, as predicted, two indicators of partner availability moderated the sexual intercourse shifts in the preregistered analyses marginally significantly: ovulatory increases were absent among women in long-distance relationships and among those who reported rarely spending the night with their partner. The daily report of whether the couple spent the night together did not moderate the shift, but the same-day behaviour may act as a mediator, not moderator, of ovulatory shifts in sexual behaviour. We see this pattern as partial support for our hypothesis 1.7., stating that ovulatory increases would be observed if circumstances allowed it, if the partner was close by. This pattern is also consistent with the findings for coupled women in a larger within-subject study on 1,180 women and 37,170 diary days (Caruso et al., 2014), but runs counter to previous results from 20,000 women in a between-subject study (Brewis & Meyer, 2005). Unexpectedly, shifts in in-pair desire also appeared to be stronger for women cohabiting with their partner.

5.1.3 Mate retention, jealousy

We observed no fertile window changes in partner mate retention, but the generalizability of change for these items was very low, making the detection of an effect unlikely. Our questions for these outcomes were based on the previous literature, in which generalizabilities of change were not reported. We had ourselves preregistered a suboptimal procedure for improving outcome reliabilities, based on assessing

Cronbach's alphas, which ignore the multilevel structure of the data. We instead calculated all analyses by item in a purely exploratory format. Based on these analyses and research published after our preregistration (Gangestad, Garver-Apgar, Cousins, & Thornhill, 2014), future research on partner mate retention should more clearly and comprehensively examine *prohibitive* behaviours, as opposed to *persuasive* behaviours, because items measuring the former seemed to show stronger changes.

5.1.4 Self-perceived desirability and clothing choices

We found fertile window increases in self-perceived desirability in our preregistered analyses that were robust to our checks, although standard errors were relatively broad because we used only a single item to assess this outcome. Contrary to our predictions, we found no fertile window changes in self-reported "sexy clothing", even though this was associated with desirability. As predicted, we also found no change in "flashy/showy" clothes and self-esteem in our robustness checks. These results are consistent with recent large-sample replications of fertile phase increases in facial attractiveness (Jones, Hahn, Fisher, Wang, Kandrik, Han, Lee, et al., 2017), suggesting that day-to-day changes in self-perceived attractiveness might track actual changes in physical attractiveness.

5.1.5 Other outcomes

For all the other outcomes we found no ovulatory changes that were also absent among HC users. Reassuringly, in no case did we observe any significant associations for outcomes for which we predicted none (relationship satisfaction, self-esteem,

spending the night/communication with the partner, female jealousy, and female mate retention). Nor did we find associations for the narcissism outcomes, for which we had indirectly extrapolated our predictions from prior reports in the literature of ovulatory changes in clothing, interpreted as signs of intrasexual competition (Durante et al., 2008). We should reiterate in this context that we did not replicate cycle shifts on clothing choices either. Perhaps this can be interpreted as evidence that the literature suffers more from potential false positives than from false negatives, though it is noteworthy that some previous studies had not found ovulatory increases in in-pair sexual desire and behaviour (Brewis & Meyer, 2005; Haselton & Gangestad, 2006). We would like to emphasize that both negative and positive results were largely robust to the many different analytic approaches that we tested.

5.2 Predicted moderator effects and individual differences

There was insufficient evidence for moderation of male mate retention behaviour, extra-pair or in-pair desire by the partner's attractiveness (no matter if assessed as relative to self, sexual, or physical), as predicted by the *good genes ovulatory shift hypothesis*. Although some patterns descriptively pointed in the predicted direction, none of the predicted patterns were significant, and some were opposite to our predictions. Because only 144 naturally cycling women remained for our preregistered analyses, statistical power may have been insufficient to detect plausible moderation effect sizes. However, we found no evidence for moderation effects in the more inclusive sample of our robustness tests. Although our sample sizes are bigger than many published studies that reported a moderation effect (Haselton & Gangestad, 2006;

Pillsworth & Haselton, 2006b), we would ideally prefer to exceed their power by a wider margin due to winner's curse, i.e. effect sizes being overestimated through selection and publication bias. We should also mention that some of the earlier studies we aimed to replicate (Haselton & Gangestad, 2006; Larson et al., 2013) did not actually report significant main effects of the fertile phase. Increases were reported to be *qualified* by a moderator. In this sense, we replicated neither findings on main nor on moderator effects from these studies. Still, we believe GGOSH can be taken to predict main effects as well, because amplified shifts in some women whose partners lack certain characteristics should, averaged across women, still yield detectable main effects. But taken literally, our findings shed doubt on GGOSH in finding no substantial moderator effects by partner attractiveness.

There are some conceptual similarities between ovulatory shift moderators of extra- and in-pair desire and direct tests of ovulatory changes in mate preferences, because both regard a shift in who is preferred as a mate. Newer, more adequately-powered laboratory research also sheds doubt on ovulatory shifts in preferences for facial masculinity (Jones, Hahn, Fisher, Wang, Kandrik, Han, Fasolt, et al., 2017) and twin studies show that heritable individual differences in this preference dwarf any cyclical changes (Zietsch, Lee, et al., 2015).

We found no evidence for the tentatively predicted moderation of increases in extra-pair desire or self-perceived desirability by neuroticism, extraversion, or shyness. However, because we had on average 25 days for each woman, we could estimate inter-individual differences in ovulatory increases (i.e. random effects for the fertile

window). Random effect variances for the fertile window predictor were substantial. Hence, there might be real heterogeneity in ovulatory increases to be explained. Future research should test and improve the reliability of these inter-individual differences across cycles. Determining if there is inter-individual variation in cycle shifts to be explained should be a precursor step before further attempts to identify both methodological and theoretically substantial moderators of ovulatory increases, such as partner or relationship attributes. Further, until any such moderation patterns are better understood, researchers should probably refrain from testing for moderation in the absence of main effects of fertility if they have not preregistered their approach, because this may lead to (accusations of) overfitting.

5.3 Theoretical implications

Although further tests should be conducted, the *good genes ovulatory shift hypothesis* could be wrong, given that we could not replicate previously reported moderators. More recent theoretical work emphasises that predictions of adaptive extra-pair sex, which (Pillsworth & Haselton, 2006a) call dual mating, should be divorced from predictions of ovulatory changes in mate preferences that do not necessarily precipitate extra-pair sex, but still function to bias sire choice (Gangestad, Thornhill, & Garver-Apgar, 2015). We cannot test all aspects of these recent theoretical developments in our study. An alternative, simpler explanation (Roney & Simmons, 2013) is based on life history theory. It suggests the observed increase in sexual desire during the fertile phase reflects a motivational priority change towards reproduction. The purported function would be to accept higher costs of sex, such as energetic and opportunity costs

or sexually transmitted infections, the more likely it is that sex leads to conception. This theory also predicts fertile phase drops in somatic investment, such as food intake (Fleischman & Fessler, 2007; Roney & Simmons, 2017). In this study, we did not assess any non-reproductive motivations, and we collected no data on single women. Hence, we cannot test whether general, target-unspecific sexual motivation drives the effects on in- and extra-pair desire we find ((Roney, 2009). Future studies should be designed and powered to discriminate between these and other theories. Relatedly, theoreticians should make exact predictions down to what certain statistical models will find, because verbal ambiguity might otherwise preclude the identification of the best supported theory.

5.4 Effect size comparison

Some perspectives (Roney & Simmons, 2013) predict a generalized increase in sex drive with fertility across the menstrual cycle, while others more specifically predict an increase in sexual interest for certain partners (Gangestad et al., 2015). These perspectives differ in predictions of whether the effect on extra-pair desire should be larger than that on in-pair desire. Although testing these competing predictions was not the goal of the present study, we can compare the relevant effect sizes. The continuous backward-counted predictor recommended in (Gangestad et al., 2016) hedges for uncertainty in the estimation of the fertile window. Our effect sizes thus account for uncertainty and reflect the estimated change when certainly in the fertile window, although the predictor never gives a more confident prediction than 58%. In Likert points from 1 to 6, the fertile window effect was 0.26 [0.17;0.35] for extra-pair desire in the

robustness check data. The in-pair desire effect had the same size: 0.26 [0.10;0.42]. We could now standardise the effects by the residual standard deviation in the multilevel model to obtain an effect size estimate of Cohen's *d*. Since the residual standard deviation of extra-pair desire is much smaller (0.61) than that of in-pair desire (1.1), their standardised estimates would differ by a factor of two. However, our items for in- and extra-pair desire were not comparable and upon inspecting item-level associations in Bayesian models that appropriately account for the ordinal nature of the Likert data, we believe comparisons between the two outcomes are futile. If we can conclude anything, effects were larger on average for items that required no object of desire to be present and no action to be taken. Future studies should attempt to settle the question of whether changes in extra- or in-pair desire are independent and different in size. Most importantly, they should test whether both can be simplified to an increase in sex drive that amplifies interest in all men without affecting their rank order, i.e. mate preferences. To do so, studies should construct parallel items to measure extra-pair, in-pair and objectless sexual desire and behaviour, and test for fertile phase changes in the rank order of ratings of male stimuli.

We suggest not to prematurely ignore the reported effects because of their small size. The effects on in-pair desire are, for instance, comparable with reported effects of a hormonal contraceptive use on sexual desire in a randomised controlled trial (Zethraeus et al., 2016). Moreover, we found evidence for substantial inter-individual variation (see below), so that effects that are small on average might be substantial for some women.

5.5 Hormonal contraception

Whenever we found an ovulatory increase, we also found that it was absent among hormonal contraceptive users. In this sense, we identified one reliable moderator. The absence of these cycle changes probably reflects the suppression of ovulation and concurrent hormonal changes. Moreover, estimated effects of menstruation and the premenstrual phase on psychological outcomes as measured in the diary were also diminished among HC users. In the preregistered analyses, we found only small and statistically non-significant mean level differences between HC users and cycling women in the diary outcomes, as well as in the demographic and personality variables that we tested. These differences are presumably confounded by selection and attrition effects. For example, women who expect their relationship to last may be more likely to start using HC and to show less extra-pair desire, and women who experience libido decreases on HC may go off it again. Thus, the (absence of) mean level differences may not (entirely or at all) speak to causal effects of HC.

There are few randomised controlled trials (RCTs) that can answer questions about psychological changes caused by HC use. Existing ones so far mostly ignore cycle phase (Zethraeus et al., 2016, 2017) thus not yielding the full picture of differences across the cycle. Potentially, this can lead to spurious or misleading conclusions of differences, if women in the naturally cycling control group are measured in different cycle phases across time points. As the effects of cycle phase on sexual desire in our study were similar in size to effects reported for hormonal contraceptives in (Zethraeus et al., 2016), further RCTs should tease cycle phase and HC influences apart.

The suppression of cyclical psychological changes is not currently being pointed out as a side effect of the pill in package leaflets, although they do mention potential effects on libido and appetite. Potentially, decreased fluctuations in extra- and in-pair desire might be seen as less worrisome than e.g. decreased average levels of libido, or altered mate preferences (Alvergne & Lummaa, 2010), but this decision is best left to HC users themselves. Decision making about HC use may vary, e.g. some women may prefer to have cyclical ups and downs, while some may prefer to have a lower but constant mean level. Moreover, individual differences in the actual physiological and psychological response to HC may be more important than differences in side effect preferences and should be a future research priority.

5.6 Limitations

In this study, we relied on self-report, which may mean that social desirability, measurement reactivity and recall error could affect our results. We hope we succeeded in minimising these issues by ensuring privacy and anonymity for participants, preventing access to past responses, asking specific closed-form questions daily, and statistically testing and adjusting for temporal trends (Barta, Tennen, & Litt, 2012). Some women in this sample may have used fertility tracking apps as a supplemental contraceptive method or simply out of interest. Such women may not have reported using these apps, because we only asked about contraception. Potentially, their increased awareness could change our results. An obvious improvement would be to also collect partner- and potentially peer-reports, although this might have negative consequences for the perceived anonymity of responses. To decrease measurement

reactivity and to test its effect, future studies could space out diary invitations over a longer period, for instance by sending them only on odd days or tailoring them to predicted (non-)fertile phases. Ideally, the schedule would be varied randomly by group (Barta et al., 2012).

We overestimated how conscientiously participants would fill out the diary. Hence, some women strung out the participation period over such a long time that menstruation could have occurred in an unobserved period, because women only reported menstrual onsets that occurred fewer than 7 days ago. Therefore, fertility was not estimable for ~6% of days (Table 3). Further, sending daily invitations via email presented a technical challenge. Due to delays in the sending process and spam filters some emails occasionally arrived a few hours late or not at all. We introduced text message reminders approximately halfway through the study and remedied this somewhat. These problems are presumably unrelated to outcomes and cycle position as *M_e5* shows, but still worth avoiding in the future. Because we required 35 complete daily reports before the study could end, some women never concluded our study, leading to 31% dropout in the follow-up survey. Future studies should use a fixed timespan for the diary, so that the follow-up takes place at the same time regardless of participation rate.

We only asked participants whether they had been intimate with someone other than their partner, but failed to systematically ask about the context and sex of the person. Free-text responses showed that several instances of reported extra-pair activity were not cheating with another man, but polyamorous or open relationships, affairs with women, or sex with the partner and another couple or a third person. All of these have

dubious relevance to the research question about adaptive benefits of extra-pair infidelity. We also did not collect data on single women, preventing us from discriminating between an increased propensity for flings in general versus extra-pair infidelity.

The generalizability of change for our outcome scales was sometimes zero and in other cases suboptimal. Previous research, from which we derived our scales, may have suffered the same problem, but did not conduct the appropriate psychometric analyses to find out. We think menstrual cycle research should learn from work on psychometrics and measurement in personality development research (Shrout & Lane, 2012). Mirroring the old person-situation debate (Kenrick & Funder, 1988), the evolutionary literature now debates the relative importance of between and within person variation (Havlíček, Cobey, Barrett, Klapilová, & Roberts, 2015; Jones, Hahn, Fisher, Wang, Kandrik, Han, Lee, et al., 2017; Zietsch, Lee, et al., 2015). However, without improving the methodology and psychometrics used to study within-person variations the debate will not be resolved (Roberts & Caspi, 2001; Shrout & Lane, 2012). Future work should also differentiate sexual activity more than we did here, including not just sexual intercourse and other sexual activity with the partner, but also masturbation and nonsexual intimacy.

Our sample was a convenience sample. Although it included a broad range of women, many (73%) were students, most (87%) had no children, few (12%) were married and all spoke German. Generalizability to older and higher-fertility populations, especially from settings that are not western, educated, industrialised, rich and

democratic (Henrich et al., 2010) may thus be limited. Although we assume universal hormonal changes drive our effects, hormonal levels might differ substantially for women who do not cycle regularly, for instance because they have recently been pregnant or breastfeeding, or because they have worse nutritional status.

Lastly, although we conducted a large number of robustness checks, we fell short of doing a full multiverse analysis or specification curve in which all possible ways to analyse the data are reported (Simonsohn, Simmons, & Nelson, 2015; Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). We decided not do this, because we believe many of our data-analytic decisions are justified properly, and multiverse analyses are most useful if no procedure was preregistered. Hence, our goal here was rather to show the effect of various approaches on the associations, as a guide to interpreting previous work as well as ours.

5.7 Suggestions for planning future and reading past cycle studies

The two most interesting takeaways from our researcher degrees of freedom simulations (see 4.1.2) might be that a) optional stopping and outcome switching had worse impacts than random covariates or switching between narrow, broad, and continuous fertile window estimates, and that b) false positives were acceptably rare (less than 5% in most conditions) if one simply applies a significance threshold of .01. The latter result only holds if researchers behaved as simulated and really stopped at p

< .05 (Nelson, Simmons, & Simonsohn, 2016), but might provide a useful guide to reading the older, non-preregistered literature.

Although it is difficult to compute an equivalent of Cohen's *d* for multilevel models, our comparable effect size estimates ranged from 0.12 to 0.43. Some were hence only a quarter of the smallest effect size (0.4) considered in (Gangestad et al., 2016). Future research would improve their odds of detecting an effect by improving the reliability of outcomes, predictors, collecting data on more women, more days, or ideally by doing all of this together. Empirically, not a single effect reported here would have been detected if we had collected only the first diary day for each woman in a between-subject analysis. Neither would effects have been detected using only two days per woman in a high vs. low fertility repeated measures design, a common design of previous studies, even though we collected ten times as many women as the average previous study. Whether the fertility predictor was formed based on forward- or backward-counting, narrow, broad, or continuous fertile phases seemed to make less of a difference (Figure 3), except that predictors using more data are preferable and that (pre-)menstruation should be adjusted for.

To fully understand the accompanying cyclical changes going along with ovulation, researchers should collect data over many days per woman (Haselton & Gangestad, 2006; Roney & Simmons, 2013). We have released our survey software and study code to make it easier to conduct online diary studies like this one (Arslan & Tata, 2016). Although online diary studies using counting methods will probably always be most cost-efficient, hormonal assays, especially repeated ones (Jones, Hahn, Fisher, Wang,

Kandrik, Han, Lee, et al., 2017; Roney & Simmons, 2013), are needed as converging evidence and to directly test hormonal mediators. They can compensate smaller affordable sample sizes through the greater validity of their predictors. Potentially the two designs can be fruitfully merged (Roney & Simmons, 2013), so that patchy hormonal assays are used to impute more valid predictors in a larger diary dataset.

Our study was preregistered. We consider this a good way in which researchers can protect themselves from unintentionally generating false positives through selection and publication bias. To combat publication bias even more effectively researchers might also try the *Registered Report* format (Nosek & Lakens, 2014) that is offered by an increasing number of journals. However, preregistration requires that the analysis procedure and ideally the data collection and cleaning procedure are set before data collection. Standard operating procedures are one way to simplify this process and to make it easier for researchers to sufficiently specify their plans, especially in areas in which they have not worked before. We have released our study materials, our data cleaning code and our code for computing menstrual onsets as potential groundwork for a standard operating procedure. We welcome improvements to this procedure that can be publicly shared. We also call for further work to improve inferences of conception probability, tailored to individual cycle lengths, regularities, and potential demographic factors. Although hormonal and sonographic measures of ovulation and fertility will retain their superiority, the use of day counting methods is justified by a much larger amount of data that can be and have already been collected efficiently (e.g. in the numerous cycle tracking apps).

Although we fail to conceive any reasonable non-hormonal or non-causal alternative explanations for the changes we observe mid-cycle, these inferences could be strengthened through a true randomised control group. We suggest that future hormonal contraceptive RCTs collect diary data across several full cycles in both experimental groups. By doing so we would be able to assess differences caused by contraceptive pills across the whole cycle, not just in e.g. the luteal phase (Zethraeus et al., 2016), and we would have sufficiently reliable within-subject data to examine heterogeneity in the response to contraceptive pills. Future studies should also attempt to better test whether awareness of being in the fertile window drives any effects.

5.8 Conclusions

In a high-powered, within-subject diary study, we were able to replicate main effects of ovulatory increases in self-perceived desirability, as well as extra-pair and in-pair sexual desire and behaviour. We failed to replicate reported ovulatory increases in partner mate retention behaviour and clothing style, and found only ambiguous support for increases in sexual behaviour. In contrast to previous reports, we found no evidence that sexual desire shifted more strongly among women who deemed their partner less sexually attractive. Previous studies had inadequate power, sometimes used suboptimal between-subject designs, and none were preregistered. Hence, several previous reports of ovulatory shifts and moderators thereof may have been false positives. We do not rule out changes along other dimensions or moderators that we and others have not tested, but large, well-designed, preregistered studies will be necessary to show these credibly. Alternatively, our data are consistent with the theory that ovulatory increases

reflect generalized changes in sexual motivation, serving the adaptive function to avoid costs associated with sex when it will not lead to conception (Roney & Simmons, 2013). Further work should directly test competing theories against each other.

6 Acknowledgements

We thank Hanne Straus for her help collecting the data on hormonal contraception, Silvia Bradatsch for extracting the sample sizes from the meta-analysis by Gildersleeve et al. (2014a), and Aileen Marske for helping with the supplement. We thank Isabelle Habedank, Maren Fußwinkel, and Sarah J. Lennartz for their helpful comments on an earlier version of this manuscript.

7 References

Alvergne, A., & Lummaa, V. (2010). Does the contraceptive pill alter mate choice in humans? *Trends in Ecology & Evolution*, *25*, 171–179. doi:10.1016/j.tree.2009.08.003

Arslan, R. C., & Penke, L. (2015). Evolutionary Genetics. In *The Handbook of Evolutionary Psychology* (Vol. 2, pp. 1047–1066). New York: Wiley. doi:10.1002/9781119125563.evpsych245

Arslan, R. C., & Tata, C. (2016). *formr.org: v0.12.6*. doi:10.5281/zenodo.60957

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*, 1531–1544. doi:10.1037/0022-3514.74.6.1531

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*, 1013–1037. doi:10.1037/a0034431

Bafumi, J., & Gelman, A. (2006). Fitting multilevel models when predictors and group effects correlate. *Social Science Research Network*. doi:10.2139/ssrn.1010095

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement reactivity in diary research. In M. R. Mehl & M. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 108–123). New York, NY: Guilford Press.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv:1506.04967 [stat]*. Retrieved from http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv:1406.5823 [stat]*. doi:arXiv:1406.5823

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Blake, K. R., Dixson, B. J. W., O'Dean, S. M., & Denson, T. F. (2016). Standardized protocols for characterizing women's fertility: A data-driven approach. *Hormones and Behavior*, *81*, 74–83. doi:10.1016/j.yhbeh.2016.03.004

Brewis, A., & Meyer, M. (2005). Demographic evidence that human ovulation is undetectable (at least in pair bonds). *Current Anthropology*, *46*, 465–471. doi:10.1086/430016

Bürkner, P.-C. (in press). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2015). Stan: A probabilistic programming language. *Journal of Statistical Software*.

Caruso, S., Agnello, C., Malandrino, C., Lo Presti, L., Cicero, C., & Cianci, S. (2014). Do hormones influence women's sex? Sexual activity over the menstrual cycle. *The Journal of Sexual Medicine*, *11*, 211–221. doi:10.1111/jsm.12348

Deschner, T., Heistermann, M., Hodges, K., & Boesch, C. (2003). Timing and probability of ovulation in relation to sex skin swelling in wild West African chimpanzees, Pan troglodytes verus. *Animal Behaviour*, *66*, 551–560. doi:10.1006/anbe.2003.2210

Dixson, A. F. (2012). *Primate sexuality: Comparative studies of the prosimians, monkeys, apes, and humans 2nd edition*. Oxford: Oxford University Press.

Durante, K. M., Li, N. P., & Haselton, M. G. (2008). Changes in women's choice of dress across the ovulatory cycle: Naturalistic and laboratory task-based evidence. *Personality & Social Psychology Bulletin*, *34*, 1451–1460. doi:10.1177/0146167208323103

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*, 891–904. doi:10.1007/s11192-011-0494-7

Fehring, R. J., Schneider, M., & Raviele, K. (2006). Variability in the phases of the menstrual cycle. *Journal of Obstetric, Gynecologic*, *35*, 376–384. doi:10.1111/j.1552-6909.2006.00051.x

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*, 120–128. doi:10.1037/a0024445

Fincham, F. D., & May, R. W. (2017). Infidelity in romantic relationships. *Current Opinion in Psychology*, *13*, 70–74. doi:10.1016/j.copsyc.2016.03.008

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*, 83–87. doi:10.1111/j.0963-7214.2004.00280.x

Fleischman, D. S., & Fessler, D. M. (2007). Differences in dietary intake as a function of sexual activity and hormonal contraception. *Evolutionary Psychology*, *5*, 642–652.

Gangestad, S. W. (2016). Comment: Wood et al.'s (2014) speculations of inappropriate research practices in ovulatory cycle studies. *Emotion Review*, *8*, 87–90. doi:10.1177/1754073915580400

Gangestad, S. W., Garver-Apgar, C. E., Cousins, A. J., & Thornhill, R. (2014). Intersexual conflict across women's ovulatory cycle. *Evolution and Human Behavior*, *35*, 302–308. doi:10.1016/j.evolhumbehav.2014.02.012

Gangestad, S. W., Haselton, M. G., Welling, L. L. M., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., … Puts, D. A. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evolution and Human Behavior*, *37*, 85–96. doi:10.1016/j.evolhumbehav.2015.09.001

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, *23*, 573–587. doi:10.1017/S0140525X0000337X

Gangestad, S. W., & Thornhill, R. (2008). Human oestrus. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 991–1000. doi:10.1098/rspb.2007.1425

Gangestad, S. W., Thornhill, R., & Garver, C. E. (2002). Changes in women's sexual interests and their partners' mate-retention tactics across the menstrual cycle: Evidence for shifting conflicts of interest. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 975–982. doi:10.1098/rspb.2001.1952

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2005). Women's sexual interests across the ovulatory cycle depend on primary partner developmental instability. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 2023–2027. doi:10.1098/rspb.2005.3112

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2015). Women's sexual interests across the ovulatory cycle. *The Handbook of Evolutionary Psychology*. doi:10.1002/9781119125563.evpsych114

Gelman, A., & Loken, E. (2014). The statistical crisis in science: data-dependent analysis — a "garden of forking paths" — explains why many statistically significant comparisons don't hold up. *American Scientist*, *102*, 460.

Gerlach, T. M., Arslan, R. C., Schultze, T., Reinhard, S. K., & Penke, L. (in press). Predictive validity and adjustment of ideal partner preferences across the transition into romantic relationships. *Journal of Personality and Social Psychology*.

Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014a). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological Bulletin*, *140*, 1205–1259. doi:10.1037/a0035438

Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014b). Meta-analyses and p-curves support robust cycle shifts in women's mate preferences: Reply to Wood and Carden (2014) and Harris, Pashler, and Mickes (2014). *Psychological Bulletin*, *140*, 1272–1280. doi:10.1037/a0037714

Harris, C. R., Chabot, A., & Mickes, L. (2013). Shifts in methodology and theory in menstrual cycle research on attraction. *Sex Roles*, *69*, 525–535. doi:10.1007/s11199-013-0302-3

Harris, C. R., Pashler, H., & Mickes, L. (2014). Elastic analysis procedures: An incurable (but preventable) problem in the fertility effect literature. Comment on

Gildersleeve, Haselton, and Fales (2014). *Psychological Bulletin*, *140*, 1260–1264. doi:10.1037/a0036478

Haselton, M. G., & Gangestad, S. W. (2006). Conditional expression of women's desires and men's mate guarding across the ovulatory cycle. *Hormones and Behavior*, *49*, 509–518. doi:10.1016/j.yhbeh.2005.10.006

Haselton, M. G., & Gildersleeve, K. (2016). Human ovulation cues. *Current Opinion in Psychology*, *7*, 120–125. doi:10.1016/j.copsyc.2015.08.020

Haselton, M. G., Mortezaie, M., Pillsworth, E. G., Bleske-Rechek, A., & Frederick, D. A. (2007). Ovulatory shifts in human female ornamentation: Near ovulation, women dress to impress. *Hormones and Behavior*, *51*, 40–45. doi:10.1016/j.yhbeh.2006.07.007

Havlíček, J., Cobey, K. D., Barrett, L., Klapilová, K., & Roberts, S. C. (2015). The spandrels of Santa Barbara? A new perspective on the peri-ovulation paradigm. *Behavioral Ecology*, *26*, 1249–1260. doi:10.1093/beheco/arv064

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*, 111–135. doi:10.1017/S0140525X10000725

Inzlicht, M., Gervais, W., & Berkman, E. (2015). Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015. *Social Science Research Network*. doi:10.2139/ssrn.2659409

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532. doi:10.1177/0956797611430953

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., … DeBruine, L. M. (2017). Within-woman hormone-attractiveness correlations are not simply byproducts of between-women hormone-attractiveness correlations. *bioRxiv*, 136515. doi:10.1101/136515

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., … DeBruine, L. M. (2017). Women's preferences for facial masculinity are not related to their hormonal status. *bioRxiv*, 136549. doi:10.1101/136549

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*, 23–34. doi:10.1037/0003-066X.43.1.23

Landolt, M. A., Lalumière, M. L., & Quinsey, V. L. (1995). Sex differences in intra-sex variations in human mating tactics: An evolutionary approach. *Ethology and Sociobiology*, *16*, 3–23. doi:10.1016/0162-3095(94)00012-V

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, *47*, 111–121. doi:10.1026//0012-1924.47.3.111

Larmuseau, M. H., Matthijs, K., & Wenseleers, T. (2016). Cuckolded fathers rare in human populations. *Trends in Ecology & Evolution*, *31*, 327–329. doi:10.1016/j.tree.2016.03.004

Larson, C. M., Haselton, M. G., Gildersleeve, K. A., & Pillsworth, E. G. (2013). Changes in women's feelings about their romantic relationships across the ovulatory cycle. *Hormones and Behavior*, *63*, 128–135. doi:10.1016/j.yhbeh.2012.10.005

Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2016, January 14). Ambitious p-hacking and p-curve 4.0. *Data Colada*. Retrieved from http://datacolada.org/45

Nosek, B. A., & Lakens, D. (2014). Registered Reports. *Social Psychology*, *45*, 137–141. doi:10.1027/1864-9335/a000192

Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke creation of synthetic data in r. *Journal of Statistical Software*, *74*, 1–26. doi:10.18637/jss.v074.i11

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. doi:10.1126/science.aac4716

Pillsworth, E. G., & Haselton, M. G. (2006a). Women's sexual strategies: The evolution of long-term bonds and extrapair sex. *Annual Review of Sex Research*, *17*, 59–100. doi:10.1080/10532528.2006.10559837

Pillsworth, E. G., & Haselton, M. G. (2006b). Male sexual attractiveness predicts differential ovulatory shifts in female extra-pair attraction and male mate retention. *Evolution and Human Behavior*, *27*, 247–258. doi:10.1016/j.evolhumbehav.2005.10.002

Pillsworth, E. G., Haselton, M. G., & Buss, D. M. (2004). Ovulatory shifts in female sexual desire. *Journal of Sex Research*, *41*, 55–65.

Revelle, W. (2017). *psych: Procedures for psychological, psychometric, and personality research*. Retrieved from https://CRAN.R-project.org/package=psych

Roberts, B. W., & Caspi, A. (2001). Personality development and the person-situation debate: It's déjà vu all over again. *Psychological Inquiry*, *12*, 104–109. doi:10.1207/S15327965PLI1202_04

Roney, J. R. (2009). The role of sex hormones in the initiation of human mating relationships. In P. T. Ellison & P. B. Gray (Eds.), *The Endocrinology of Social Relationships* (pp. 246–269). Cambridge: Harvard University Press.

Roney, J. R., & Simmons, Z. L. (2013). Hormonal predictors of sexual motivation in natural menstrual cycles. *Hormones and Behavior*, *63*, 636–645. doi:10.1016/j.yhbeh.2013.02.013

Roney, J. R., & Simmons, Z. L. (2017). Ovarian hormone fluctuations predict within-cycle shifts in women's food intake. *Hormones and Behavior*, *90*, 8–14. doi:10.1016/j.yhbeh.2017.01.009

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 1913–1917. doi:10.1098/rspb.1999.0866

Schilling, K. M., Straus, H., Arslan, R. C., Gerlach, T. M., & Penke, L. (2014). Cycle shifts and pill dosage effects. Retrieved from osf.io/98w3h

Schwarz, S., & Hassebrauck, M. (2008). Self-perceived and observed variations in women's attractiveness throughout the menstrual cycle—a diary study. *Evolution and Human Behavior*, *29*, 282–288. doi:10.1016/j.evolhumbehav.2008.02.003

Shrout, P., & Lane, S. P. (2012). Psychometrics. In *Handbook of research methods for studying daily life*. Guilford Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *Social Science Research Network*. doi:10.2139/ssrn.2694998

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*, 160384. doi:10.1098/rsos.160384

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702–712. doi:10.1177/1745691616658637

Stirnemann, J. J., Samson, A., Bernard, J.-P., & Thalabard, J.-C. (2013). Day-specific probabilities of conception in fertile cycles resulting in spontaneous pregnancies. *Human Reproduction*, *28*, 1110–1116. doi:10.1093/humrep/des449

van Elk, M., Matzke, D., Gronau, Q., Guang, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: A

skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, 1365. doi:10.3389/fpsyg.2015.01365

Wegienka, G., & Baird, D. D. (2005). A comparison of recalled date of last menstrual period with prospectively recorded dates. *Journal of Women's Health*, *14*, 248–252. doi:10.1089/jwh.2005.14.248

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *22*, 392–399. doi:10.2307/2346786

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

Wood, W., Kressel, L., Joshi, P. D., & Louie, B. (2014). Meta-analysis of menstrual cycle effects on women's mate preferences. *Emotion Review*. doi:10.1177/1754073914523073

Zethraeus, N., Dreber, A., Ranehill, E., Blomberg, L., Labrie, F., Schoultz, B. von, … 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*, 1238–1245. doi:10.1016/j.fertnstert.2017.02.120

Zethraeus, N., Dreber, A., Ranehill, E., Blomberg, L., Labrie, F., von Schoultz, B., … Hirschberg, A. L. (2016). Combined oral contraceptives and sexual function in women - a double-blind, randomized, placebo-controlled trial. *The Journal of*

Clinical Endocrinology and Metabolism, *101*, 4046–4053. doi:10.1210/jc.2016-2032

Zietsch, B. P., Lee, A. J., Sherlock, J. M., & Jern, P. (2015). Variation in women's preferences regarding male facial masculinity is better explained by genetic differences than by previously identified context-dependent effects. *Psychological Science*, *26*, 1440–1448. doi:10.1177/0956797615591770

Zietsch, B. P., Westberg, L., Santtila, P., & Jern, P. (2015). Genetic analysis of human extrapair mating: heritability, between-sex correlation, and receptor genes for vasopressin and oxytocin. *Evolution and Human Behavior*, *36*, 130–136. doi:10.1016/j.evolhumbehav.2014.10.001

Appendix D.
Curriculum vitae

Curriculum vitae omitted from online version