Effects of policy on fertility: A systematic review of (quasi)experiments*

Janna Bergsvik†, Agnes Fauske‡ and Rannveig K. Hart§

Draft submitted for consideration at the European Population Conference 2020

PLEASE DO NOT CITE

Abstract

This paper describes the results of a systematic review of the literature of policy effects on fertility in Europe, USA, Canada and Australia. Empirical studies were selected through extensive systematic searches, with subsequent literature list screening. Inclusion was conditional on implementing an experimental or quasi-experimental design. A total of 59 published papers and recent working papers were included, covering the topics of parental leave, childcare, health services, universal child transfers and welfare reforms. Childcare and universal transfers seem to have the most positive effects on fertility. Few effects were found for parental leave, but this could be linked to these reforms not lending themselves to efficient (quasi)experimental evaluation. Withdrawing cash transfers to families through welfare reforms has limited fertility effects. Subsidizing assisted reproductive technologies show some promise in increasing birth rates of women above age 35.

* We are grateful to Øystein Kravdal for comments on an earlier version of the paper and to librarian Gyri Straumann at the Department of Reviews and Health Technology Assessments, Norwegian Institute of Public Health, for conducing the basic search, and for guidance on databases and searching tools when conducting later searches. We acknowledge funding from the "Determinants of falling fertility" project, funded by several Norwegian Ministries.

† Sociodemographic Research Group, Research Department, Statistics Norway. Email: jbk@ssb.no

‡ Department of Health and Inequality, Norwegian Institute of Public Health. Email: agnes.fauske@gmail.com

§ Department of Health and Inequality, Norwegian Institute of Public Health. Email: Rannveigkaldager.hart@fhi.no

1. Introduction

The decline of fertility below replacement levels has been met with concern in several advanced economies (McDonald, 2006). In 2017, 83 of 201 countries in the world had fertility below replacement levels (United Nations, 2018). At the same time, many of these countries allocate large budget shares to family support in different forms. In 2015, 66 percent of the European governments and almost 40 percent of Asian governments had policies to raise fertility or at least impede further decline (United Nations, 2018). Within Europe, cross-country studies show that extensive public support to families correlates with higher fertility (see e.g. Gauthier and Hatzius (1997), Kalwij (2010) and Wood, Neels, and Vergauwen (2016)). Seminal studies assessing within-country change over time find that fertility trends often follow policy change closely (Hoem, 1990; Rønsen & Skrede, 2010). However, strong and stable economic conditions facilitate costly policies and may themselves contribute to relatively high fertility, questioning whether family policies indeed are the key driver.

A small but growing literature of (quasi-)experimental studies tries to isolate fertility effects of specific policies. While quasi-experimental studies aim at finding universal effects of single policies and bear lower risk of interpreting other societal changes as policy effects, they constitute a «lower bound» for actual policy effects because spillovers induced by the policy usually are not captured (see also (Olivetti & Petrongolo, 2017). Results from these studies have not yet been targeted in an updated and systematic review. This literature review fills this gap and synthesizes studies that take an experimental or quasi-experimental approach in studying the effect of policy on fertility. For external validity, our review is limited to countries within Europe, USA, Canada and Australia. Hence, our systematic review complements a large literature of comparisons between these countries and across time, deepening our understanding of the interplay between public policy and fertility decisions.

2. Theoretical starting point

Raising children takes time and money, and public policies can influence fertility by affecting these resources. In its simplest form, the economic theory of the family postulates that the number of children a couple chooses to have depends on the availability of time and money, as well as their preferences for spending that time and money on children or other purposes (Becker, 1991). Policies as cash transfers to families, tax breaks for parents, subsidized childcare and parental leave directly affect parents' time and money budget through increasing family income or reducing the direct or indirect (opportunity) cost of children. But also, policies not directly targeting families affect family resources and the cost of children, e.g. housing and health care subsidies. If children are a "normal good" (i.e. a good for which consumption increases in income), increased income or decreased costs will translate into larger family sizes. However, several mechanisms make the expected relationship between resources and fertility more complicated.

First, having more resources could make parents wish or feel obliged to invest more in each child, e.g. provide better housing or schooling. This would in turn increase the cost of raising a child and could reduce the demand for children. Such a quality-quantity trade-off (and the preference for quality above quantity) can lead to (counterintuitive) negative income effects (Becker, 1991).

Further, the expected relationship between resources and fertility is complicated by the fact that (at least one parent in) most families earn the bulk of their income in the labour market, and that several family benefits such as tax breaks for parents and most parental leave benefits depend on active employment. With increasing wages and stronger attachment to the labour market, the losses from taking time off work to care for children (the opportunity cost) increase, too. While this *substitution* effect complicates a precise understanding of the fertility effect of employment related benefits, it is not evoked by all policies. Unconditional cash transfers do

not invoke a substitution effect. In contrast, tax breaks for parents and several parental leave benefits typically strengthen parents (i.e. mothers) labour attachment and could invoke the substitution effect. Their expected effect on fertility is hence more theoretically ambiguous.

In addition, one should expect substantial variation in policy responses in different population groups. A simple example is that reduced kindergarten fees relieve the family budget and reduce the price of future children, but not for families with a strong preference for parental care over formal care. Similarly, cash transfers conditional on not using formal care constitute an income/price effect for families positive to home care but should not influence families with a firm preference for formal care.[5] Parental leave benefits reduce substitution costs in the first period of childrearing, if they compensate income losses from taking time off work to care for children. However, the policy is less relevant for not eligible families, e.g. when the main carer already is out of paid work, or for parents with a strong preference for paid over unpaid work, because they would return to work quickly regardless of compensation.

Finally, policies may influence fertility by redistributing the time cost of childbearing between the parents. Opportunity costs have been disproportionally taken by mothers, and if this has dampened fertility, policies aiming to shift opportunity costs to fathers may have pro-natalist effects. However, such effects will emerge only if father's increased opportunity costs not negatively impact fertility.

Last, fertility is also influenced by norms, fecundity and regulation costs (Crimmins, 1985). Politically influencing norms and preferences is typically seen as both difficult and, in liberal democracies, largely undue (Schultz, 2015). Hence, policies that affect fertility will typically work through affecting the time and money available to parents. In this literature review, we include relevant policies regardless of their aim, be it fertility increase, welfare-to-work-

[5] Except indirectly: if the demand for formal care exceeds coverage and the policy frees places.

initiatives or simply cutbacks driven by budgets deficits. We note that policies may have an explicit pro-natalist (or anti-natalist) intent, and that these intentions may have effects in their own right. For instance, a welfare reform designed to reduce nonmarital childbearing while on welfare sends a strong signal that this is unwanted behaviour, while a "baby bonus" rewards childbearing.

Of the policies we consider, some types of health services, such as health services for children or perinatal care, work through income and price effects and are theoretically akin to transfers. Other types of health services directly influence regulation costs. More specifically, when contraception and abortion is cheap and accessible, the cost of preventing unwanted pregnancies falls, and so should fertility. Our starting point is, however, that fertility is a private choice that is enabled or constrained by the context provided by public policies. Policies that use restrictions of elective abortion as a means of fertility increase will not be considered. We will not consider the literature on availability of contraception specifically but consider fertility effects when the cost of contraception is changed as part of a package of changing costs of health services.

3. Evaluating effects of policies on fertility

Identifying the causal effect of policies on fertility requires research designs that overcome problems of selection (selective implementation and/or uptake), omitted-variable bias (confounding unmeasured factors) and multicollinearity (masking the relationship between the policy and fertility). How different model specifications can lead to contradictory conclusions is demonstrated for example in Rindfuss et al. (2007; 2010), showing that a naïve estimation of the association between childcare availability and fertility provides a negative relation between these two at the municipal level in Norway. Conversely, a specification that accounts for the non-random distribution of childcare facilities across the country shows the expected positive

effects. Probably, public childcare expanded faster in areas where women's work-family conflict was most pressing, and where fertility initially was lower. If one is interested in the causal effect of providing public childcare on these women's fertility it is crucial to use a good counterfactual – how would their fertility have looked if there was no/less/more public childcare?

The studies included in this review use (more or less) formalized strategies to tackle the above-mentioned identification challenges. They use advanced panel data models, experiments, or analytical design exploiting reforms as natural experiments to get good comparison groups. This part briefly introduces how effects of policies on fertility are measured in the included studies and discusses some general traits of the different strategies, i.e. experimental studies, regression discontinuity designs, difference-in-differences analyses and fixed effects panel regression (see also Angrist and Pischke, 2009).

Randomized experiments where a benefit is randomly given to some persons (treatment group) and not to others (control group), provide the most obvious opportunity for evaluating the causal effect of that benefit. However, for practical and ethical reasons experiments are rare, and external validity may be limited if experiments create superficial settings.

Regression Discontinuity-designs (RD) use naturally occurring random variation in treatment eligibility. They are suitable when arbitrary cut-offs define who is affected by a policy change. In the included studies, most often the birthdate of a child defines whether old/new legislation applies. If the cut-off indeed is set arbitrary and if it is not possible for parents to select into treatment status (e.g. to time delivery), those being just eligible should be similar to those being just ineligible and provide good comparison groups. Rigorous tests and placebo analyses usually come with credible RD studies.

Difference-in-differences estimation (DiD) builds on the same logic. Some units (individuals or municipalities) are exposed to policy changes or expansions, while others are

not. Within-unit fertility changes over time are then compared between the units to see if the trends developed substantially different among those affected by a policy. A causal interpretation is given based on the assumption of parallel trends absent the policy change. Sensitivity tests, i.e. placebo-analyses, are again much used to strengthen credibility and show the plausibility of underlying assumptions. In cases where confounding trends are identified triple-differenced designs are also used *(DiDiD).*

Two-way fixed effects panel regression models (2W FE) are a generalized form of difference-in-differences estimation. By using time and unit fixed effects these models effectively control for confounding time shocks and time constant differences between units. They provide causal estimates if no unmeasured time-varying variables bias the results. Credible studies provide sensitivity tests, while placebo-analyses mostly are not feasible

It makes sense to reflect on how quasi-experimental study designs define treatment and control groups. Who is affected by a policy and who remains unaffected? First, some policies create persisting differences in available family resources between population subgroups. For example, when higher monthly cash transfers are given to eligible families but not to others, treatment and control groups are easily identified, and the challenge is to prove that they are identical on other characteristics. Other reforms are universal and create only temporary differences (around the implementation period) between treated and untreated families. For example, in regression discontinuity designs extensions of (universal) parental leave are frequently evaluated based on eligibility differences imposed by reform implementation dates and one child's birthdate to define treatment and control groups. In most cases, parents in the control and treatment group both would receive longer leave for the next child. Hence, the incentives for continued childbearing are identical.

In this case another analytical distinction is useful to bear in mind. Policies can reduce existing costs of children already conceived (current child effect) and/or lower the anticipated

cost of future children (future child effects). Some studies of the fertility effects of policies distinguish between these current and future child effects, also referred to as income and price effects. If such a distinction is possible depends on the nature and time-horizon of the policy reform and the analytical design used to evaluate their effect. If applicable, it is expected that temporary differences between treatment and control groups in experiences with current children (induced for example by the implementation date of a parental leave reform) will have least influence on long-term fertility differences between these two groups. If a policy can be evaluated based on comparing variation on persistently available resources (e.g. childcare availability or cash transfers) or future child effects (giving incentives to some parents but not others), its potential short- *and* long-term effects will be more easily detectable.

Typically, policy effects on fertility outcomes are measured both in the short- (e.g. timing of first births or spacing to the next birth), as well as in the long-run (e.g. number of children several years after a reform or completed fertility). In cases where reforms only invoke short-term differences between comparison groups, timing effects are more easily detected than quantum effects. However, having children earlier might also raise the total number of children, because more fecund years remain for subsequent births. But fertility might as well reverse again after immediate reform effects.

Last, a comparison of fertility between treatment and control groups seldomly recognizes social interaction effects, which may impact policy take up as well as fertility behaviour also beyond the directly affected population. Such interdependencies may affect fertility outcomes of the control group, and comparing fertility responses between treated and 'untreated' parents would underestimate the policy effect (see (Olivetti & Petrongolo, 2017). Social interaction effects have for example been shown in take-up of parental leave both for mothers (Welteke & Wrohlich, 2019) and fathers (Gordon B. Dahl, Løken, & Mogstad, 2014).

4. Methods

This section briefly describes the search and selection process, and how narrative synthesis is used to synthesize the results. Details can be found in the protocol (Bergsvik, 2019). The project is also pre-registered at PROSPERO (R. K. Hart, Bergsvik, J. and Fauske, A., 2019).

4.1. The process of search and sorting

The *bibliographic database search* was carried out using relevant social and medical science databases (Epistemonikos, Social services abstracts, Cochrane library, Medline, Web of science, Popline, Sociological abstracts, RePec). The original search string constrained outcomes to various measures of fertility, and, for larger databases, constrained methods to those with potential for causal inference. No constraints were set for the explanatory variable (intervention). In a next step, the original search string was modified and extended with key words for two types of interventions, family policy and housing. Together, these searches generated 17 228 unique hits.

Empirical studies were included if they fulfilled the *selection criteria* regarding participants, intervention, comparison, outcomes and study design (PICOS) (Liberati et al., 2009). Our criteria are described in Table 1, with further details in the protocol. In addition, recent review articles were included for literature list screening if they reviewed articles that matched our PICOS criteria well.

A PRISMA diagram (Liberati et al., 2009; Moher D, 2009) documenting the *screening process* is found in Figure 1. Titles and abstracts from the 17 228 articles found through the systematic search were screened for relevance and method by two researchers, using the web application Rayyan (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016). When studies were included for full text reading, reference lists were screened for relevant articles, that again were read in full text by two researchers. A total of 13 review articles was also screened (Balbo,

Billari, & Mills, 2013; Blank, 2002; Gauthier, 2007; Hantrais, 1997; Lichter & Jayakody, 2002; Lopoo & Raissian, 2012; Mills et al., 2011; Neyer & Andersson, 2008; Olivetti & Petrongolo, 2017; Pirog & Ziol-Guest, 2006; Tach & Edin, 2017; Thévenon & Gauthier, 2011; Thévenon & Luci, 2012). In sum, 332 articles were read in full text by two researchers, of which 59 constitute the final sample.

Two researchers were involved in evaluating the risk of bias in the studies included, with a third to resolve disagreement. *Bias assessment* was done by evaluating the extent to which assignment was (conditionally) random (quasi-random, quasi-experimental), and by evaluating tests for conditional randomness (see Angrist and Pischke (2008) for an overview of relevant tests). Results were considered more credible and given greater weight in the narrative synthesis if robustness checks were done for fertility outcomes specifically and linked to the subgroup/outcome where a significant effect (if any) was found.

4.2. Narrative synthesis

Our analysis of the material is a narrative synthesis guided by the four steps developed by Popay (2006), see also Ryan (2016). Section 5 gives a detailed description of each of the included studies in terms of both text and overview tables, structured by type of intervention. The discussion in Section 6 focuses on patterns in data, in terms of intervention type, evaluation design, context and subgroups. We also critically assess the completeness of evidence, and variation in this across type of intervention, as well as our applied methods for evaluation of bias (i.e. the validity of the identification strategies).

5. Description of patterns by topic

5.1. Parental leave

Parental leave gives parents (mothers) the right to take time off from work in relation to a birth and new-born care while being granted to return to the pre-birth job afterwards. Job-protected

parental leave comes unpaid, state-paid and employer-paid, and can fully or partly compensate for income losses during the absence. Long-term costs, for example in the form of a worse income development after the absence might remain. The extent of and eligibility criteria for parental leave compensations vary considerably between countries and/or states, often they depend on mothers' employment status or earnings prior to the birth. In addition, many countries also reserve some weeks for fathers.

Parental leave reforms lead to plenty policy variation over time within countries. Such reforms or in one case the introduction of parental leave itself have been used in all 11 studies that were included in this review. Three studies examine effects of general parental leave expansions, while three studies examine effects of introducing or expanding the paternity leave. Two studies look at fertility effects of going from means tested to earnings related maternity leave benefits. Studies are summarized in Table 2.

Parental leave was introduced in the United States in 1993 through the Family and Medical Leave Act (FMLA). FMLA provided 12 weeks unpaid job-protected leave to employees with stable employment at a covered employer over the previous 12 months. Cannonier (2014) compares fertility trends between women fulfilling the eligibility criteria and not eligible women in a difference-in-differences design and finds an increased probability of having a first and second birth, as well as earlier births among eligible women after the introduction of FMLA.

The Nordic countries were among the first to implement extensive parental leave schemes, and four studies in this review examine reform effects of these. G. B. Dahl, Loken, Mogstad, and Salvanes (2016) use samples of mothers giving birth around the implementation dates of six parental leave expansions in Norway between 1987-1992 in a regression discontinuity design to study reform effects on completed fertility. They only find a small effect

of the 1992 reform on the number of children born to mothers 14 years after and conclude that, overall, the expansions of paid leave did little to encourage fertility.

Focusing on a Swedish parental leave reform from 1989, Liu and Skans (2010) investigate effects of prolonged parental leave on the timing and number of future children. The reform retroactively prolonged the leave period for parents with a birth in 1988/89 from 12 to 15 months. Using a difference-in-differences approach they find a small increase (0.24 percentage points for an additional month of leave entitlement) in the probability of having an additional child within 18 months of the last, which appears to be driven by highly educated mothers. No significant effect was found for the total number of children.

In 1993 Norway introduced a four week «father's quota» in the parental leave scheme. A share of the parental leave period was from then on reserved for fathers. Importantly, the 1993 reform extended the total parental leave length. To be eligible for the father's weeks both parents had to fulfil the eligibility criteria for parental leave benefits. Eligibility requires employment in 6 of the last 10 months and income above a certain ceiling. Cools, Fiva, and Kirkeboen (2015) use a difference-in-differences approach to compare the effects of this extension to the 1992 parental leave extension, which came without reserving a share for fathers. They find no effects of paternity leave on parents' child spacing or total number of children 14 years after the reform.

In 2009 the share of reserved parental leave for fathers was extended from 6 to 10 weeks in Norway. Fertility effects of this extension are studied by R. K. Hart, Andersen, and Drange (2019). Results from their regression discontinuity analysis show no effects on progressions to further children within 5 years.

Spain introduced two weeks of paid paternity leave in 2007. Farre and Gonzalez (2018) examine the fertility effects of paternity leave in Spain using both regression discontinuity and

difference-in-differences strategies, finding that the probability of having another child within 6 years was lower and the spacing was longer among eligible couples. Results were mostly driven by mothers over 30.

Using a difference-in-differences design Cygan-Rehm (2016) examines effects of a German maternity leave reform on West German mothers' subsequent fertility within 12/ 21/ 24/ 33/ 36/ 45/ 48/ 57 months. The 2007 reform made maternity-leave benefits earnings related instead of means tested and changed the length of the grace period, securing eligibility for benefits when having a next child within a short time after the focal child (also 'speed premium'). Cygan-Rehm finds that the reform significantly affected the timing of higher-order births in line with heterogeneous economic incentives given by the reform. Negative and persisting effects on the probability of having another child were found for the lowest-income mothers. In contrast, for mothers who were 'reform winners' relatively weak and temporary positive effects on higher-order births were found.

The same German reform was used by Raute (2019) to compare fertility responses of high earning women, defined as the most significantly treated by the reform, to those of low earning women in a difference-in-differences analysis. Raute's study is (together with Cannonier 2014 and Ang 2015) one of the few parental leave studies included in this review that examine effects also on first births, finding that after the reform the highly educated were more likely to have a first and second child.

In Austria parental leave comes with a flat rate benefit. A reform in 1990 increased the leave period from 12 to 24 months and prolonged the 'speed premium' for the next child. In 1996 the speed premium as well as the leave period were shortened again to 18 months of parental leave. Lalive and Zweimuller (2009) study effects of these reforms on higher order births using a regression discontinuity design finding that extending parental leave with one year gives about 12 additional children per 100 women. Following the reduction in 1996 they

find compressed spacing between first and second births but no effect on the number of second births within three years.

In 2006 the Quebec Parental Insurance Program (QPIP) increased the generosity of parental leave benefits in Quebec through increasing the maximum insurable earnings and the income replacement rate from 55 to 70 percent for 30 out of 55 weeks of the leave period. Using a difference-in-differences strategy Ang (2015) finds that this program increased birth rates by 23.5 percent compared to other Canadian provinces. Effects were particularly strong for first and second parity.

In line with the diversity and complexity of parental leave policies, the corresponding fertility effects are highly dependent on the population under scrutiny, the extent of the studied reforms and consequently the differences that reforms actually create between treatment and control groups (as discussed in chapter 3). While half of the studies identified various timing effects after parental leave reforms, the effects on completed fertility are more ambiguous. No general effects of the parental or paternity leave extensions were found in the Nordic countries (Liu and Skans 2010; Dahl et al. 2016; Cools et al. 2015; Hart et al. 2019), and in Spain the introduction of paternity leave even had negative effects (Farre and Ganzalez, 2018).

Reforms that altered benefits substantially gave more apparent fertility effects (e.g. Lalive and Zweimuller 2009; Ang 2015). Further, studies comparing fertility effects between eligible versus not eligible groups (in the long run) find positive effects on first and second births among eligible women (see Raute (2019), Cannonier (2014) and Ang (2015)). Highly educated women respond better to earnings-related parental leave benefits (Raute, 2019; Cygan-Rehm, 2016; Liu and Skans, 2010).

5.2.Childcare

Access to childcare can reduce the conflict of work and family responsibilities for parents. Hence, childcare availability, affordability and acceptance are strongly linked to the opportunity costs of childrearing. Childcare can be offered by relatives, bought in the private market or provided publicly. The extent to which these different options are used and available varies considerably between countries, and access to publicly provided childcare, especially for the youngest children, has expanded relatively recent and could in theory influence fertility.

In this review seven studies were included for childcare, summarized in Table 4. Four studies use variation in the availability of local childcare centres over time between municipalities/ counties,[6] while two studies focus on reforms changing the costs of publicly provided childcare. One study uses pension reforms to examine how reduced availability of grandparental care impacts the fertility of the offspring of the generation affected by delayed retirement.

Rindfuss et al. examine the effect of childcare availability on first birth timing (2007) and completed fertility (2010) for the cohorts of mothers born in Norway 1957-1962. Both studies employ two-way fixed effects and use variation in the percentage of pre-school-age children in childcare centres within municipalities between the years 1973 and 1998. For first birth timing they find that increased childcare availability relates to an earlier transition to motherhood as well as higher probabilities of becoming a mother at every age for the age groups 15-19, 20-24, 25-29, 30-35. Rindfuss et al. (2010) extends the focus to total number of children born to women by age 35. They find an increase of slightly more than 0.1 in average number of children born for each 10 percentage points increase in childcare availability. The increase

[6] Studies relying on a combination of time and *region* fixed effects to identify the effect of childcare on fertility are not included in this overview. Region fixed effects are considered too broad to fully capture the endogeneity of variation in childcare center placements at the municipality level (as shown in Rindfuss et al. 2007 compared and referring to Kravdal 1996).

is significant and positive for all parities, albeit the largest absolute difference is found for transitions to second births, and the largest relative difference for third births.

Wood and Neels (2019) estimate the effect of local childcare coverage in Belgium on the probability of having a child between 2002 to 2005 for the population of dual-earner couples in 2001.[7] The study uses municipality fixed effects and variation in the number of childcare places over the population aged 0-3 from 2002 to 2005. Changes in childcare coverage within a municipality are positively associated with birth hazards. Because the study doesn't include time fixed effects, common time trends might bias the estimates.

Bauernschuster, Hener and Rainer (2016) study the effect of local childcare coverage in West German counties on birth rates among women aged 15-44. Using public childcare slots over the population of children under the age of three from 1998 to 2010 they study the fertility effect of several reforms (2005-2008) that led to a large-scale staggered expansion of public childcare for those children. First, in a difference-in-differences framework births per 1000 women are compared between counties with above-median and below-median childcare increases between 2002 and 2009. Then, a continuous measure provides effects using the full variation in childcare availability. Results show that the provision of public childcare has positive effects on fertility. A 10 percentage point increase led to an increase in birth rates of 2.8 percent. Effects were driven by married women and were strongest for second and third births.

Using a difference-in-differences framework Gathmann and Sass (2018) examine effects of the 2006 «home care subsidy»-reform in the East German state of Thuringia. The reform effectively raised the price of choosing public childcare compared to home care. It transferred at least 150 Euros monthly to those not sending their 2-year-old child to public

7 The dual-earner criteria probably samples a broader and more representative population for first birth probabilities than for second or third births. Because of this restriction, subsample results comparing findings by parity are not commented.

childcare. In a first step, fertility responses of families with 2-year-old-children in Thuringia versus other East German states before and after the reform were compared. Among them, those with 2 or more children showed small positive effects. These families were 'reform winners' because the subsidy was increasing with number of children. Further, fertility effects were stronger for single, low-income, and foreign parents. In a second step, the fertility effect for all women aged 18-45 in Thuringia was estimated, finding that the home care subsidy discouraged first births and had negligible effects on families with children.

Mork, Sjøgren and Svaleryd (2013) focus on the fertility effect of a Swedish childcare reform announced in 1998, implementing a user fee cap in 2002. The reform standardized childcare fees across municipalities and imposed a maximum fee cap, which for most families reduced childcare costs. However, new charges per child were dependent on household income and age and number of children. Thus, costs were reduced more for some families than others. Using a difference-in-differences design Mork et al. (2013) compare before and after reform fertility at the household type x municipality level over the years 1996-2003. Among married couples an early positive effect on first births is observed. Their fertility increased by 9.8 percent, primarily driven by low-income households. Second births were postponed, and higher order births increased with 14.5 percent, but these last effects were only marginally significant.

Finally, Battistin, De Nadai and Padula (2015) use several Italian pension reforms between 1992 and 2001 that delayed retirement ages to estimate the effect of grandparental availability on their offspring's fertility. Results from the regression discontinuity analyses show that reduced availability of grandparents has negative effects on fertility, and that it is much so in families where family ties are stronger. Family ties are measured by an index using several variables about both partners' relationship with the family of origin (i.e. distance, support, contact). Formal childcare availability somewhat attenuates these effects, especially where family ties are weak.

To sum up, results are in line with expectations. Increasing childcare availability has positive effects on fertility (Rindfuss et al. 2007; Wood & Neels 2019), especially on higher order births (Rindfuss et al. 2010; Bauernschuster et al. 2016). Parents are those directly experiencing the benefits of available childcare, while childcare availability might not be as salient for those not yet having a child. In the same line has reduced availability of grandparents negative effects in a context where grandparental care is an important option (Battistin et al. 2015). Changes in the price of childcare also affect subgroups of couples in line with theoretical expectations. Increasing the price of choosing public childcare compared to home care discouraged first births but increased the fertility of those known to be more prone to choose home care, i.e. single, low-income, or foreign parents and those with many children (Gathmann and Sass 2018). Contrary, lowering and standardizing the prices of public childcare as in the Swedish case had positive effects on fertility, particularly on first births and interestingly already after announcement (before implementation) (Mork et al. 2013).

Findings on parity specific responses to changes in childcare availability and prices diverge to some degree. While childcare availability had stronger impact on the fertility of those who already were parents (Rindfuss et al. 2010; Bauernschuster et al. 2016), reforms on the price of public childcare seem to move parents only marginally. Instead reform effects emerged for first births, and one could speculate whether the diverging results can be explained by 'announcement' effects of childcare reforms on those who are not yet parents, while actual availability (without announcement) has stronger effects on those experiencing the gains.

5.3. Health services

Perinatal care and health services for the new child constitute a large proportion of immediate costs of having a child. In extensive welfare states, this cost is carried collectively rather than individually, and will hence not influence fertility choices directly. The cost of health services may impact fertility through three main mechanisms. First, among parents, reducing the cost of

health care for children already born constitutes a transfer and generates a positive income effect. Second, reduced costs of prenatal and perinatal care, as well as health services for children, lower the price of the next child. For subfecund couples, reduced costs of reproductive technologies will have a similar price effect. Thirdly, and in contrast, reducing the cost of contraception and abortion reduces regulation costs, potentially inhibiting fertility – particularly in age groups where births tend to be unplanned or unwanted. In sum, reducing the cost of health services have potentially ambiguous fertility effects. Effects are expected to depend both on the type of services provided and the age of the affected population, with younger women more likely to use health services to limit fertility.

11 studies were included for health services, all based on data from the USA. Studies are summarized in Table 5. Eight studies look at variation in the cost of all health services, stemming from an experiment with free insurance coverage (one study), changes in Medicaid (four studies), the Affordable Care Act (ACA, two studies), and a health care reform in Massachusetts (one study). Three studies look at the effect of reducing the cost of infertility treatment specifically.

Leibowitz (1990) analyses fertility effects of a health insurance experiment carried out in six US cities 1974-1979. In the experiment families were randomly assigned to different insurance schemes, including a fully covered plan (i.e. free health services) for up to five years. Free health services lower the cost of inhibiting conception, as well as the cost of pregnancy, delivery and childrearing. Birth rates were 29 per cent higher among fully covered women than in the control group, an effect that emerged after two-three years. The study cannot conclude whether completed fertility is affected, or births are simply shifted to a period where health services are cheaper (Leibowitz, 1990, p. 709).

Medicaid provides health insurance to women and families with low income and covers a large share of the costs of perinatal care, delivery and health services to children. In the 1980s

and 1990s, the eligibility threshold for families with children (including childless but pregnant women) has been expanded multiple times, with substantial variation in timing and level across states (Deleire, Lopoo, & Simon, 2011). All four studies on Medicaid effects utilize a state and year fixed effects design and a cell-based estimation strategy, where birth rates are calculated separately by race, educational attainment, and marital status.

The earliest Medicaid study by T. Joyce, Kaestner, and Kwan (1998), estimates the effects of two Medicaid expansions on fertility rates, finding positive effects on birth rates. The subsequent studies all use more refined and more plausibly exogenous measures of Medicaid availability, a simulated fraction of women eligible (Deleire et al., 2011; Zavodny & Bitler, 2010) or an expansion threshold (Lincoln H. Groves, Hamersma, & Lopoo, 2018; Zavodny & Bitler, 2010). Deleire et al. (2011) and Zavodny and Bitler (2010) both take log quarterly birth rates as their outcome. While Zavodny and Bitler (2010) find a positive effect among women with lower education, Deleire et al. (2011) find no robust effects after detailed controls for demographic characteristics. Lincoln H. Groves et al. (2018) analyse first, second and higher order births separately, and find significant positive effects on higher-order births, concentrated among women with high school education only.

Two studies analyse an aspect of The Affordable Care Act ("Obamacare") implemented in 2010, where dependents up to age 26 could be listed on their parent's employer insurance. This reduced the cost of contraception and birth/perinatal care among a large share of young adults. Both Abramowitz (2018) and Heim, Lurie, and Simon (2018) use a difference-in-differences design, with unaffected age groups as controls. Both find negative effects on birth rates. Abramowitz (2018) find (non-significant) indications that increased use of hormonal contraceptives may mediate this effect, while abortion rates are unmoved. Heim et al. (2018) find indications that those not enrolled in post-secondary education drive the effects. Both studies show robustness checks in the form of pre-trend tests.

Apostolova-Mihaylova and Yelowitz (2018) utilizes a state-specific expansion of health insurance in Massachusetts in 2006, using neighbouring states as controls in a difference-in-differences design. The reform reduced the cost of all health services and has been seen as a predecessor to the ACA reform. They find an 8% reduction of fertility among unmarried women aged 20-34, where births are often unplanned. Among married women in the same age group, fertility increases by 1%, an unsurprising response to lowering the cost of births in a group where fertility intentions are high.

Compared to lowering the cost of all health services, lowering the cost of infertility should have more unambiguous fertility effects. Infertility treatment lowers the cost of having children despite fecundity problems, and should increase birth rates among the sub-fecund, who are overrepresented at higher ages. Effectively, cheaper infertility treatment lowers the cost of fertility postponement, potentially causing age at first birth to increase. US states have discretion to allow or require that employer mandated insurance covers infertility treatment, and to specify the types of insurance schemes and infertility treatments to be included. This generates variation in the price of infertility treatment across space and time in the US, and the three included studies utilize this variation to estimate effects of the cost of fertility treatment on fertility in variations of difference-in-differences designs.

L. Schmidt (2005) finds a 32% increase in first birth rates among women above age 35, concentrated among whites. L. Schmidt (2007) expands on this finding, showing that effects are larger when a larger population is covered, and finds no effects at higher order births. Machado and Sanz-de-Galdeano (2015) utilize the same variation to estimate effects on age at first birth as well as completed fertility, finding that cheaper fertility treatment leads to postponed first birth, with no effects on completed fertility. Machado and Sanz-de-Galdeano (2015) uses a synthetic control approach in addition to a standard DiD-design and offer extensive visual displays of pre-trends. Still, long-term effects on timing of births and

completed fertility are inherently difficult to measure in most quasi-experimental designs, warranting some caution in the interpretation of results. The combination of a postponement effect at low ages and a positive effect above age 35 is consistent with theoretical expectations. There is some tension between a positive effect above 35, driven by couples who would otherwise have struggled to bear children, and no effect on completed fertility.[8] This conflict suggests that further research is required before strong conclusions on the effect on completed fertility can be drawn.

The empirical findings confirm that reducing the cost of health services has ambiguous effects on fertility. Among young adults, results from the ACA reform indicate that fertility is lowered when health services are cheaper (Abramowitz, 2018; Heim et al., 2018), perhaps due to more consistent contraceptive use (Heim et al., 2018). However, Apostolova-Mihaylova and Yelowitz (2018) find that a similar reform in Massachusetts increased fertility for married women up to their mid-30s. Among women above age 35, positive fertility effects emerge when infertility treatment is cheaper (L. Schmidt, 2005; L. Schmidt, 2007), though these effects may be temporary (Machado & Sanz-de-Galdeano, 2015). A general reduction in the cost of health services in all age groups, as induced by Medicaid expansions, seems to have a weak positive effect on fertility among high school educated women (Deleire et al., 2011; Zavodny & Bitler, 2010), concentrated at higher parities (Lincoln H. Groves et al., 2018). Five years of free health care has substantial positive fertility effects, yet these are likely to be at least temporary (Leibowitz, 1990). Despite some conflicting evidence, reducing the price of infertility treatment at higher ages stands out as the most effective strategy to increase birth counts. For welfare states that already offer comprehensive free or low-cost health services, expanding access to infertility services shows some promise in stimulating birth rates.

[8] Mechanically, this would require a negative effect below age 35, potentially because some couples postpone childbearing due to better insurance and then adapt to a child free lifestyle and remain childless.

5.4. Universal child transfers

Cash transfers to families with children raise the family income and reduce the costs of current and future children (income and price effect) and should consequently have positive effects n fertility. However, two factors might dampen these positive effects. First, parents may use additional transfers to invest more in children already born (i.e. substitute quality for quantity). The presence of such effects is illustrated by studies showing that transfers improve child health (REF). Second, if transfers are given as tax breaks, they will also invoke a negative substitution effect, potentially lowering fertility (see Section 2). A large empirical literature on the effect of tax breaks on labour supply (REFS) illustrate the plausibility of such a substitution effect.

This review includes eight studies on fertility effects of universal and unconditional cash transfers and tax breaks based on policy changes in European contexts (Spain, Germany, Norway) or other extensive welfare regimes (Canada), summarized in Table 6. Targeted and conditional transfers are summarized in the next subchapter.

Four studies analyse transfer expansions specific to the Canadian province Quebec, using (parts of) the rest of Canada as controls. Milligan (2005) analyses the effect of an increase in cash transfers to families with children in 1988. The increase was particularly marked for third children, and he finds strong effects at third births in a difference-in-differences design. He presents extensive robustness checks in terms of pre-trend tests, but data limitations hinder him from fully addressing these. Ang (2015) analyses the same reform (among others), using better micro data. Contrary to the reform incentives, she finds effects concentrated at first birth. However, this latter paper provides only limited robustness checks and pre-trend inspections. Also Parent and Wang (2007) analyse effects of the same reform using better data on cohort fertility. While replicating Milligan (2005)'s result for immediate effects, they conclude that no effects on cohort fertility persist. Analysing Quebec-specific extensions of parental allowance in the 1970, Kim (2014) reaches similar conclusions about effects emerging in the short term

but waning in the long term. Overall, the evidence from Canada points toward marked, yet transitory, effects of universal cash transfers on fertility. The many region-specific policy changes in Quebec complicate identification of the precise effect of each policy. While each of the included studies aim to address this, we note that the challenge will necessarily be larger when addressing long term than short term effects.

Two of the included studies on universal transfers are based on reforms in Spain. González (2013) analyses the effect of a one-time cash payment ("baby bonus") introduced to all Spanish residents in July 2007. The immediate implementation and sharp cut-off of this reform makes it well suited for a regression discontinuity (RD) design. She finds a statistically significant increase in conceptions following the reform, as well as a (somewhat smaller) significant decrease in abortion rates. Azmat and Gonzalez (2010) evaluate the effect of a 2003 reform of the Spanish income tax, aiming to increase fertility while upholding maternal labour supply. The reform introduced substantial tax breaks for households with young children, and additional deductions conditional on mothers working. They find a 5% increase in fertility, combined with increased labour supply of mothers of small children. It should be noted that while the labour supply models are standard difference-in-differences models, the fertility models have no clear control group and pertain to first difference or regression discontinuity designs. While the 2007 reform allows for stronger causal identification, the evidence taken together indicates that monetary incentives have pro-natalist effects in the Spanish context.

Riphahn and Wiynck (2017) study the effect of a German child benefit reform in 1996. The reform, and hence the identification of effects, is complicated: In general, first births got better subsidized for lower educated (lower earning) couples, while second births were better subsidized for the higher educated (higher earning). The authors use these educational differences for identification in a difference-in-differences design. Consistent with theory, they find positive, but only moderately robust, effects on higher order births among higher educated

couples. For first births, an unexpected negative effect emerges for lower educated couples, perhaps attributable to compositional effects with respect to age and geography. The concentration of effects among higher-earning couples is consistent with the findings of Milligan (2005) for Canada. In sum, this study (weakly) supports a positive effect of transfers on fertility.

Galloway and Hart (2015) analyse the effect of increased cash transfers and tax breaks to families with children in Norway. They exploit variation from a regional reform in a difference-in-differences design, using bordering municipalities as controls. The combination of an increased cash transfer to parents, and general tax breaks (largest for mothers with no coresidential partner), gives an increase in nonmarital first births. The result is robust to specification tests and trend modelling.

Taken together, studies of universal monetary transfers indicate a positive effect on fertility, confirmed by studies from Canada, Spain, Germany and Norway. However, long term fertility effects tend to be more difficult to study in a quasi-experimental design. The limited number of existing studies on completed fertility indicate that effects are transitory. It is important to note that these studies typically change the *current* cash flow to parents (with one-time baby bonuses as an exception) in addition to reducing the price of the marginal child. Comparing intervention and control groups where the *current and future* economic situations differ should make it easier to capture reform effects.

5.5. Welfare reforms

Welfare reforms analysed in the newer quasi-experimental literature are typically intended to strengthen labour market attachment among welfare recipients and reduce their reliance on cash transfers. Interventions are typically bundled in "packages", where job training and work incentives, reduced cash transfers and increased tax breaks are crucial elements.

Simultaneously, tax breaks for low-income working (parents), such as the US Earned Income Tax Credit (EITC) are expanded. Such credits could induce either a (positive) income effect or a (negative) substitution effect on fertility. In the US, cash transfers to unmarried women (Aid to Families with Dependent Children, AFDC) have been suspected to increase nonmarital fertility. Cuts in the AFDC, often termed "family caps", explicitly aim to reduce nonmarital childbearing. A change in (nonmarital) fertility due to such complex reforms could be due to normative pressure (as nonmarital childbearing while on welfare is directly portrayed as an outcome to be avoided), increased substitution costs due to a stronger labour force attachment, or an income/price effect from changes in cash transfers. While the net effect of a reform often is of great interest to policy makers, it is often less informative about theoretical mechanisms.

The studied welfare reforms have predominantly been implemented in the US and the UK, both liberal welfare regimes (Esping-Andersen 1990). This is also reflected in the origin of studies included (18 from the US, two from the UK). One study analyses a German labour market reform with similarities to the welfare packages from the US and UK. Studies that evaluate child support enforcement are excluded, as this in practice works more as a means to increase men's cost of a nonmarital birth (with potential negative effects on nonmarital fertility) rather than as a transfer to unmarried women (Garfinkel, Huang, McLanahan, & Gaylin, 2003). Further, a large earlier literature on variation in welfare benefits largely relies on methods that are not strictly quasi experimental, and interested readers are referred to the succinct summary by Moffit (1998). The 21 included studies are summarized in Table 7.

Two of the included studies examine effects of Aid to Families with Dependent Children (AFDC) utilizing over time within state variation in benefit size in two-way fixed effects designs. Hoffman and Foster (2000) find positive effects on nonmarital births, yet these are strongly sensitive to specification. Robins and Fronstin (1996) find positive effects on nonmarital fertility for non-white women without a high school degree.

Eleven studies assess the "family cap", which denies further AFDC cash assistance to (higher order) children conceived when their mother is on welfare, potentially reducing non-martial fertility. Five studies are based on field experiments. Four of these analyse an experiment in New Jersey 1992-1997, where a random sample of about 8 300 participants were randomly assigned to either treatment (new restrictions) or control (no changes). In addition to reducing cash transfers, the program also incentivized job training.

Jagannathan and Camasso (2003) and Jagannathan, Camasso, and Killingsworth (2004) both conclude that the program reduced fertility among blacks with limited prior welfare experience who lived in predominantly non-black neighbourhoods. M. J. J. C. E. P. Camasso (2004) found that both cash transfer reductions and job training had independent negative effects on fertility. However, Jagannathan, Camasso, and Harvey (2010) concluded that only 2.5% of the effect of the reform package could be attributed to changes in cash transfers, indicating that changes in norms towards childbearing on welfare was an important component of the reform effect. Despite some general criticisms of this experiment (Dyer & Fairlie, 2004), it gives the most reliable estimates of the effect of this particular combination of reforms. Fein (2001) analyses a comparable experiment implemented in Delaware 1995-1996, and found no significant effects on fertility.

Five studies use between-state, across time variation to study effects of family caps in two-way fixed effects designs. Three studies use a state-year fixed effects model with controls for other welfare changes, and find a negative effect on nonmarital fertility among blacks (M. J. Camasso & Jagannathan, 2009, 2016; Sabia, 2008). Horvath-rose, Peters, and Sabia (2008) find negative effects on nonmarital births, but *positive* effects on marital births. They conclude that the positive effects on marital fertility are too strong to be a response to family cap incentives, and rather indicate that state implementation of family caps is endogenous to fertility

trends. Kearney (2004) focuses on higher-order births, as only these are directly affected by the family cap, finding no effects.

A crucial challenge for identification is that family caps may be endogenous to fertility trends (e.g. states implement family caps as a response to increases in nonmarital fertility), and that their implementation may correlate with changes in other welfare schemes. More rigorous identification strategies further question the validity of the state-year fixed effects for analysing effects of the family cap. Difference-in-differences designs allow for more careful comparison of trend and control states, and the construction of more plausible control groups.

Using double- and triple-differenced designs, neither Ted Joyce, Kaestner, Sanders, and Henshaw (2004) nor Dyer and Fairlie (2004) find robust effects of family caps on fertility. Grogger and Bronars (2001) analyse variation in welfare benefits stemming from twin births, comparing duration to the next birth following a twin birth in high-welfare relative to low-welfare states. While lower benefits for children already born slows down parity progression among blacks, variation in benefits for the potential next (marginal) child, as induced by the family cap, has no effect on fertility. In sum, these more rigorous studies do not support that capping transfers to unmarried mothers on welfare limits fertility. To the extent that experimental evidence indicates effects, these effects are largely (97.5%) due non-monetary mechanisms.

Four studies assess the effects of tax breaks for working parents at the lower end of the income distribution, two from the US (EITC) and two from the UK. In the US, EITC seems to have a negative effect on white women's fertility, both on first (Reagan Baughman & Dickert-Conlin, 2003) and higher order births (R. Baughman & Dickert-Conlin, 2009). For blacks both positive effects on first births (Reagan Baughman & Dickert-Conlin, 2003) and no robust effects emerge (R. Baughman & Dickert-Conlin, 2009). Potentially, better labour market opportunities among white women invoke stronger substitution effects. A tax break similar to

the EITC was implemented in the UK in 1999. Francesconi and van der Klaauw (2007) find an no significant effect on lone mother's parity progression in a difference-in-differences design. Brewer, Ratcliffe, and Smith (2012) find that the reform increased fertility among coupled women, where the (partner's) income effect will dominate the (woman's) substitution effect.

In sum, it seems that targeted tax breaks have ambiguous fertility effects, and that the net effect will be positive only if the substitution effect is relatively weak (women has poor labour market opportunities) or the income effect is strong (works through partner's earnings). This result stands in contrast to those obtained for universal tax reductions which significantly increased fertility (e.g. in Spain shown by Azmat and Gonzalez (2010), previous chapter). Suggesting that high-income families, not eligible for targeted credits, have a more positive fertility response to tax breaks. At the same time, studies of the family cap indicate that the intention of the policy matters, as effects might be mediated by normative responses. Hence, when the intervention primarily aims at increased labour supply, this may limit positive fertility effects.

Hofmann and Hohmeyer (2013) use the announcement of a 2003 welfare reform in Germany as an instrument for economic uncertainty. The reform tightened unemployment benefits and provided conditions of economic activity comparable to those seen in welfare programs in the US and UK. They explore if fertility behaviour changes in the period between announcement and implementation, a period in which they show that perceived economic uncertainty increased. Other period changes are an obvious threat to identification, but this threat is alleviated by placebo and robustness tests. Instrumented this way, economic uncertainty reduces fertility only when perceived by the woman, and only for higher-order births. Despite identification challenges, results illustrate that economic reforms may influence fertility beyond their effect on resources available to the family.

Finally, Lincoln H Groves, Lopoo, and Issues (2018) investigate fertility effects of US federal aid to students with one deceased or disabled parent in a difference-in-differences design. The program provided substantial financial aid to students up to age 22 conditional on being unmarried. An untreated, post-program group is available as the program was revoked in the mid-80s. Insignificant DiD-estimates when control variables are outcomes support a valid identification strategy. Results show no effects on completed fertility, but an increase in age at first birth. The latter outcome is conditioned on ever having children (i.e. endogenously conditioned), which could compromise causal identification. While this specific program is not targeted towards fertility behaviour, the results indicate that students respond to economic incentives when it comes to fertility timing, perhaps leaving some room for incentivizing earlier parenthood among students. We see this reform as relevant not to welfare reforms specifically, but rather to how student's economic conditions shape their family formation behaviour.

The discussed welfare reforms typically aim to strengthen labour supply, with reduced (nonmarital) fertility as a more or less intended side effect. While theory predicts that fertility may be influenced, effects are in practice very limited. Stronger labour market attachment is sometimes achieved, and sometimes translates to lower fertility through a stronger substitution effect. Withdrawing cash transfers seem to have very limited effects on fertility, and non-monetary channels seem paramount for any effects found.

The contrast between the findings in the welfare literature and the literature on universal transfers indicates both that different population groups respond differently to monetary incentives, and, potentially, that the *intention* of the policy matters. Pro-natalist policies might show stronger effects when these intentions are clearly announced. On the other hand, announcing policies that increase economic uncertainty may have the opposite effect.

5.6. Housing

House prices might affect fertility in two different directions. First, housing is a major cost associated with family increases. High housing prices might supress fertility through increasing the costs of having a(nother) child for those who would need more living space. At the same time, for homeowners an increase in house prices implies an increase in wealth. This could have positive effects on their fertility. Hence, the effect of housing prices on fertility differs for renters/owners and by current dwelling size. But housing prices do also reflect the general prosperity of an area which could have effects on fertility independent of own wealth or the costs of living space. There are several ways through which policies affect and regulate the real estate market. However, in all studies which came across our search, variations in the cost of housing come from variations in real estate market prices over time and between areas. These are mostly not resulting of policies and hence outside the scope of our review.

6. Discussion

6.1. Patterns by study type

In this review, study type refers to two distinct yet interrelated elements: The nature of the intervention, and the nature of the evaluation design.

Patterns by intervention

The interventions we study are extremely varied. But based on the extensive summaries above, some general patterns emerge. Three groups of policies tend to impact fertility positively. First, increased availability and reduced cost of childcare both have positive fertility effects. Second, lowering health care costs may have some general positive effects on fertility through lowering the cost of children. Most importantly, however, reducing the cost of assisted reproduction has a positive impact on fertility in age groups where subfecundity is high. Third, universal transfers to families with children tend to increase fertility, even if they are given as tax breaks. All these three groups of policies are efficiently evaluated.

For two groups of policies, few effects on fertility are found. First, various reductions in welfare payments, predominantly in the US and UK, seem to have very small or no effects on fertility. The design of these policies allows for rigorous evaluations of income and price effects. The absence of effects hence indicate that no effects exist.

A second group of policies that, somewhat more surprising, yield few effects, is parental leave. Long compensated parental leaves constitute very large transfers to parents, and their (yearly) value will often largely exceed the value of e.g. kindergarten subsidies. The absence of measured effects might be linked to two characteristics of the reforms themselves. Most importantly, for reforms on universal parental leave the difference between the control and treatment group tends to lie in temporary experiences with current child(ren) – i.e. the treatment group has had a slightly longer or better compensated leave, or a different division between mother and father. (While, strictly speaking, better compensated parental leave gives current child income effects, this effect will typically be temporary and not affect the persisting flow of income when another child is considered.) These types of current child effects are a priori least likely to impact future fertility (see also chapter 2). Second, even in countries with very long compensated parental leaves, increases (and hence quasi-experimental evaluations) typically happen in relatively small increments. Even if a year of compensated parental leave impacts fertility, effects of smaller increments may be too small to be detected even with large data sets. Hence, our observation is that the available designs leave us with insufficient information to conclude.

Patterns by identification strategy

A recurrent feature in our material is that different studies analyse the same reform with different designs and conclusions, providing an excellent window to compare design and identification strategies. Prominent examples of this include changes over time in cash transfers

in Quebec, recent German parental leave reforms and between-state, between year variation in "family caps" in the US.[9]

In general, "stricter" or more conservative identification strategies will be less likely to yield results biased away from zero, and indicate effects where there are none. The patterns in our empirical material confirm this expectation. For instance, the family cap literature illustrates that the (less conservative) two-way fixed effects are more likely to yield significant reform effects than the (more conservative) double- or triple-differenced designs. The latter designs more efficiently investigate and net out deviating trends across treatment and control groups.

Running field experiments on policies is expensive and to some extent politically controversial, and our evidence leans heavily towards quasi-experimental evaluations. The experiments we include are exclusively from the US, and regard welfare reforms and health insurance expansions. It is noteworthy that in both cases, the estimated experiment effects tend to be larger than effects of comparable policies isolated in quasi-experimental designs. Various explanations could be offered for these differences. First, experiments may entail data sources with less measurement error (which would give less bias towards the null). Second, there are multiple pieces of evidence of "announcement effects" or normative effects, and it is possible that these are particularly strong when field experiments are implemented. Finally, experiments tend to be time limited. As such, individuals in the treatment group may display strong but temporary responses to what they expect to be temporary changes in the costs of childbearing.

6.2. Effect variations by outcome

The studies we look at differ in whether the outcome is measured at the aggregated or individual level. We do not observe systematically different results based on this distinction and note that

[9] Unfortunately, one may suspect that papers that slightly modify the design of a previous study and get a very similar result are less likely to be published or even submitted. The potential of such publication bias makes it difficult to assess the extent to which reanalysis tend to alter results.

the distinction between the data types is not very sharp as data can be aggregated by a very large number of categories, and individual level data will eventually be aggregated to group means in a regression analysis.

Whether the outcome is parity-specific or for all births does matter for results, but we consider this to be an issue of subgroup estimation rather than outcome measurement.

In general, it is easier to detect tempo(rary) effects than effects on completed fertility. Usually measurement error is larger for completed fertility, biasing results toward zero. Changing policies or moving in and out of policy areas/eligibility also means that reform effects will more often be "washed out" for completed fertility. A conclusion of previous reviews has been that tempo is more easily influenced than quantum (Gauthier 2007). While our review does not counter this conclusion, we would like to add that when specific reforms are considered, tempo effects are easier to identify than quantum effects. We are hence reluctant to conclude that public policy does not matter for the level of completed fertility. While the timing of birth is of importance for future population structure, completed fertility is even more crucial. Clearly, this question warrants further research.

6.3. Effect variations by subgroup

Regarding cash transfers and tax breaks, the most obvious prediction from theory would be that the largest effects would be found in lowest income brackets. In this group, the relative size of a flat transfer will be larger. Empirical studies tend to, if anything, show the opposite pattern, with larger effects in the highest income brackets (e.g. Milligan 2005), and generally weaker effects of welfare policy changes, which specifically target low income groups.

In line with theory do policies such as parental leave and public childcare that aim at enabling mothers labour force participation, have stronger fertility effects on higher educated mothers or those with a strong attachment to the labour force.

[Parity specific variation will be added]

Studies of health services display strong subsample effects, largely consistent with expectations. Young adults are more likely to put cheaper health services toward contraception (lowering births), while lower cost of ART yield positive effects only at higher ages where subfecundity is common.

In general, studies that estimate mean population effects only (e.g. Dahl et al, 2016) are less likely to find effects than studies that also look at subsamples. Sometimes, the nature of subsample estimations is obvious from the design of the policy: The family cap yields income effects for unmarried mothers only, and the strongest effects should be expected in this group. Expectations of subsample effects can also be derived from theory, but it is noteworthy that findings do not always confirm to expectations.

A challenge with the subsample estimation, given that results often do not confirm to theoretical expectations, is that of multiple testing. The mean population effect is estimated with one statistical test. Tests by four dichotomized stratifying variables (such as the commonly used ethnic background, education, age and marital status) gives 16 statistical tests, a level at which one will often, in the long run, see at least one false positive result in each study with a 5 percent significance level. Given that pre-registration is extremely rare in this literature, it is difficult to know whether the choice of stratification variables is derived from theory or post-hoc motivated after extensive testing. A stronger tradition for pre-registration would strengthen the credibility of this literature.

6.4. Effect variations by context and completeness of evidence

In general, our review has revealed a consistency in evidence across contexts. This holds especially regarding results for subgroups, e.g. positive effects of parental leave for highly educated women's fertility in the US, Canada, Sweden and Germany. However, there is also a

tendency of similar reforms being implemented in similar contexts. Very generally speaking, universal transfers, kindergarten expansions and parental leave compensations tend to take place in already relatively extensive welfare states in Europe and Canada, where they tend to positively impact fertility, while more rudimentary welfare states such as the US tend to implement cutback reforms. Hence, we have limited evidence on how extensions would work in rudimentary welfare states, and how cutbacks would influence fertility in extensive welfare states.

For health services, our empirical evidence is entirely from the US, with potentially limited validity in European welfare regimes. The federal US system, with substantial regional policy discretion, simply provides more opportunities for quasi-experimental evaluation than many European welfare states characterized by nationwide rights and reforms. Despite this skewedness of evidence, we consider the literature on health services to be informative also for the European context. Studies that look at specific health services such as assisted reproductive technologies are relevant for ongoing European discussions on the extent to which such treatments are to be publicly funded.

6.5. Methods of bias minimization

In this review, we have assessed study quality based on the criteria for a valid (quasi-)experiment established in this research literature. We have set criteria for bias minimization in the pre-registration document at PROSPERO and elaborated on them in the research protocol. While we are aware of more formal strategies developed by Cochrane (REF), we find these to only fit modestly well with an evaluation of the quasi-experimental literature. Compared to a classic Cochrane review, our synthesis leans more heavily on expert judgement rather than on pre-defined criteria.

On the other hand, we have more formalistic and pre-defined criteria for bias assessment than is common in literature reviews within the field of demography. We believe that this has added some structure and replicability to our review, perhaps also facilitating a subsequent debate with counterarguments to our judgements. While it would be of interest to develop more rigorous criteria for the assessment and synthesis of quasi-experimental studies, this has been outside the scope of our study.

7. Conclusion

In this article, we have summarized studies of the effect of policies on fertility, based on an extensive and systematic search of both published articles (> 17 000 screened) and working papers. We have found five groups of policies that are evaluated with respect to whether they influence fertility: Parental leave, childcare, universal transfers, health services and welfare. Of these, especially childcare, universal transfers and some types of health services tend to have positive effects on fertility.

Concerns about falling fertility are mostly linked to concerns about declines in future labour supply, and countries who aim to increase fertility tend to simultaneously want to preserve or even increase maternal labour supply (ref). Which of the evaluated policies unite these goals? The most obvious policy is accessible and reasonably priced childcare. In contexts where childcare coverage is high, one can speculate that improvements on accessibility, such as opening hours compatible with non-standard work hours, or quality could have further positive effects. We do not have empirical studies that assess these dimensions, however.

Universal transfers also seem to increase fertility, at least when they are substantial. A challenge with large universal transfers is that they may act as a disincentive to paid work, particularly among women with many children and relatively low human capital. However, when transfers are given as tax breaks, one largely works around this dilemma.

Finally, a targeted intervention that may have a small but significant effect is offering subsidized assisted reproduction treatment at all ages where the success rate is of meaningful size. Finding the efficiency cut point for provision would require a cost-benefit analysis, but that such services increase birth rates if available at a low cost to couples likely to benefit from them is founded in data.

Second, what does *not* work? The welfare cutbacks seen in the last decades in the US and UK seem to have very limited impact on fertility. Whether such policy packages in "reversed form" have the potential to *increase* fertility is dubious, and a reform package that increases fertility while reducing labour supply at the lower end of the income distribution also does not seem politically feasible. Given the importance of family income for child health and wellbeing (ref), such an attempt would also raise some ethical concerns.

Our largest knowledge gap seems to be on the effect of parental leave. This is not due to a lack of studies, but rather because the nature of parental leave reforms make them difficult to evaluate. Given that long parental leaves are costly, evaluating them in an experimental design akin to the US tradition would provide important insights. It would, however, only be politically feasible to randomly allocate *additional* parental leave benefits or rights, meaning that effects would be evaluated at yet another margin.

Finally, we note that several studies point towards announcement effects of policies. Policies probably perceived as supportive by parents (to be) tend to show less ambiguous fertility effects compared to policies with mixed signals. Policies with the main aim to increase (female) labour supply will naturally show heterogeneous effects because they target the family life of different population groups in different ways and aim to achieve goals that sometimes conflict with high fertility.

References [not complete]

Abramowitz, J. J. J. o. P. E. (2018). Planning parenthood: the Affordable Care Act young adult provision and pathways to fertility. *31*(4), 1097-1123.

Ang, X. L. (2015). The Effects of Cash Transfer Fertility Incentives and Parental Leave Benefits on Fertility and Labor Supply: Evidence from Two Natural Experiments. *Journal of Family and Economic Issues, 36*(2), 263-288. doi:10.1007/s10834-014-9394-3

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*: Princeton university press.

Apostolova-Mihaylova, M., & Yelowitz, A. (2018). HEALTH INSURANCE, FERTILITY, AND THE WANTEDNESS OF PREGNANCIES: EVIDENCE FROM MASSACHUSETTS. *36*(1), 59-72. doi:10.1111/coep.12235

Azmat, G., & Gonzalez, L. (2010). Targeting fertility and female participation through the income tax. *Labour Economics, 17*(3), 487-502. doi:10.1016/j.labeco.2009.09.006

Balbo, N., Billari, F. C., & Mills, M. (2013). Fertility in Advanced Societies: A Review of Research. *European Journal of Population, 29*(1), 1-38. doi:http://dx.doi.org/10.1007/s10680-012-9277-y

Baughman, R., & Dickert-Conlin, S. (2009). The earned income tax credit and fertility. *Journal of Population Economics, 22*(3), 537-563. doi:10.1007/s00148-007-0177-0

Baughman, R., & Dickert-Conlin, S. J. A. E. R. (2003). Did expanding the eitc promote motherhood? , *93*(2), 247-251.

Becker, G. S. (1991). *A Treatise on the Family*: Harvard university press.

Bergsvik, J., Fauske, A. and Hart, R.K. . (2019). *Protocol for "Effects of policy effects on fertility: A systematic review of (quasi)experiments"*.

Blank, R. M. (2002). Evaluating Welfare Reform in the United States. *Journal of Economic Literature, 40*(4), 1105-1166. Retrieved from http://www.jstor.org/stable/3217318

Brewer, M., Ratcliffe, A., & Smith, S. (2012). Does welfare reform affect fertility? Evidence from the UK. *Journal of Population Economics, 25*(1), 245-266. doi:10.1007/s00148-010-0332-x

Camasso, M. J., & Jagannathan, R. (2009). How Family Caps Work: Evidence from a National Study. *Social Service Review, 83*(3), 389-428. doi:10.1086/648190

Camasso, M. J., & Jagannathan, R. (2016). The Future of the Family Cap: Fertility Effects 18 Years Post-Implementation. *Social Service Review, 90*(2), 264-304. doi:10.1086/687368

Camasso, M. J. J. C. E. P. (2004). Isolating the family cap effect on fertility behavior: Evidence from New Jersey's family development program experiment. *22*(4), 453-467.

Cannonier, C. (2014). Does the Family and Medical Leave Act (FMLA) Increase Fertility Behavior? *Journal of Labor Research, 35*(2), 105-132. doi:10.1007/s12122-014-9181-9

Cools, S., Fiva, J. H., & Kirkeboen, L. J. (2015). Causal Effects of Paternity Leave on Children and Parents. *Scandinavian Journal of Economics, 117*(3), 801-828. doi:10.1111/sjoe.12113

Crimmins, E. M. (1985). *The Fertility Revolution: A Supply-Demand Analysis*: Chicago: University of Chicago Press.

Cygan-Rehm, K. (2016). Parental leave benefit and differential fertility responses: evidence from a German reform. *Journal of Population Economics, 29*(1), 73-103. doi:10.1007/s00148-015-0562-z

Dahl, G. B., Loken, K. V., Mogstad, M., & Salvanes, K. V. (2016). What Is the Case for Paid Maternity Leave? *Review of Economics and Statistics, 98*(4), 655-670. doi:10.1162/REST_a_00602

Dahl, G. B., Løken, K. V., & Mogstad, M. (2014). Peer Effects in Program Participation. *American Economic Review, 104*(7), 2049-2074. doi:10.1257/aer.104.7.2049

Deleire, T., Lopoo, L. M., & Simon, K. I. (2011). Medicaid Expansions and Fertility in the United States. *Demography, 48*(2), 725-747. doi:http://dx.doi.org/10.1007/s13524-011-0031-6

Dyer, W. T., & Fairlie, R. W. (2004). Do family caps reduce out-of-wedlock births? Evidence from Arkansas, Georgia, Indiana, New Jersey and Virginia. *Population Research and Policy Review, 23*, 441-473.

Farre, L., & Gonzalez, L. (2018). *Does Paternity Leave Reduce Fertility?* Retrieved from https://EconPapers.repec.org/RePEc:iza:izadps:dp12023

Fein, D. J. (2001). Will welfare reform influence marriage and fertility? Early evidence from the ABC demonstration. *J Evaluation Program Planning, 24*(4), 427-444.

Francesconi, M., & van der Klaauw, W. (2007). The socioeconomic consequences of "in-work" benefit reform for British lone mothers. *Journal of Human Resources, 42*(1), 1-31. Retrieved from <Go to ISI>://WOS:000243888300001

Garfinkel, I., Huang, C. C., McLanahan, S. S., & Gaylin, D. S. (2003). The roles of child support enforcement and welfare in non-marital childbearing. *Journal of Population Economics, 16*(1), 55-70. doi:10.1007/s001480100108

Gauthier, A. H. (2007). The impact of family policies on fertility in industrialized countries: a review of the literature. *Population Research and Policy Review, 26*(3), 323-346. doi:10.1007/s11113-007-9033-x

Gauthier, A. H., & Hatzius, J. (1997). Family benefits and fertility: An econometric analysis. *Population Studies-a Journal of Demography, 51*(3), 295-+. doi:10.1080/0032472031000150066

González, L. J. A. E. J. E. P. (2013). The effect of a universal child benefit on conceptions, abortions, and early maternal labor supply. *5*(3), 160-188.

Grogger, J., & Bronars, S. G. J. J. o. P. E. (2001). The effect of welfare payments on the marriage and fertility behavior of unwed mothers: Results from a twins experiment. *109*(3), 529-545.

Groves, L. H., Hamersma, S., & Lopoo, L. M. (2018). Pregnancy Medicaid Expansions and Fertility: Differentiating Between the Intensive and Extensive Margins. *Population Research and Policy Review, 37*(3), 461-484. doi:http://dx.doi.org/10.1007/s11113-018-9465-5

Groves, L. H., Lopoo, L. M. J. J. o. F., & Issues, E. (2018). Federal Financial Aid and Family Formation: Examining the Social Security Student Benefit Program. *39*(3), 436-444.

Hantrais, L. (1997). Exploring Relationships between Social Policy and Changing Family Forms within the European Union. *European Journal of Population, 13*(4), 339-379. doi:http://dx.doi.org/10.1023/A:1005941907983

Hart, R. K., Andersen, S. N., & Drange, N. (2019). *Effects of extended paternity leave on union stability and fertility*. Retrieved from https://EconPapers.repec.org/RePEc:ssb:dispap:899

Hart, R. K., Bergsvik, J. and Fauske, A. (2019). A systematic review of the causal effects of public policies on birth rates in Europe, Northern America and Australia. PROSPERO 2019 CRD42019134807 Retrieved from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019134807

Heim, B., Lurie, I., & Simon, K. (2018). The Impact of the Affordable Care Act Young Adult Provision on Childbearing: Evidence From Tax Data. *Demography, 55*(4), 1233-1243. doi:http://dx.doi.org/10.1007/s13524-018-0692-5

Hoem, J. M. (1990). Social policy and recent fertility change in Sweden. *Population and Development Review, 16*(4), 735-748.

Hoffman, S. D., & Foster, E. M. J. J. o. H. R. (2000). AFDC benefits and nonmarital births to young women. 376-391.

Hofmann, B., & Hohmeyer, K. (2013). Perceived Economic Uncertainty and Fertility: Evidence From a Labor Market Reform. *Journal of Marriage and Family, 75*(2), 503-521. doi:http://dx.doi.org/10.1111/jomf.12011

Horvath-rose, A. E., Peters, H. E., & Sabia, J. J. (2008). Capping Kids: The Family Cap and Nonmarital Childbearing. *Population Research and Policy Review, 27*(2), 119-138. doi:http://dx.doi.org/10.1007/s11113-008-9076-7

Jagannathan, R., & Camasso, M. J. (2003). Family cap and nonmarital fertility: The racial conditioning of policy effects. *Journal of Marriage and Family, 65*(1), 52-71. doi:http://dx.doi.org/10.1111/j.1741-3737.2003.00052.x

Jagannathan, R., Camasso, M. J., & Harvey, C. J. J. o. S. S. R. (2010). The Price Effects of Family Caps on Fertility Decisions of Poor Women. *36*(4), 346-361.

Jagannathan, R., Camasso, M. J., & Killingsworth, M. J. J. o. L. E. (2004). New Jersey's family cap experiment: do fertility impacts differ by racial density? *, 22*(2), 431-460.

Joyce, T., Kaestner, R., & Kwan, F. (1998). Is Medicaid pronatalist? The effect of eligibility expansions on abortions and births. *Family Planning Perspectives, 30*(3), 108-113, 127.

Joyce, T., Kaestner, R., Sanders, K., & Henshaw, S. (2004). Family cap provisions and changes in births and abortions. *Population Research and Policy Review, 23*(5-6), 475-511. doi:http://dx.doi.org/10.1007/s11113-004-3461-7

Kalwij, A. (2010). The impact of family policy expenditure on fertility in western Europe. *Demography, 47*(2), 503-519. doi:10.1353/dem.0.0104

Kearney, M. S. J. J. o. H. R. (2004). Is there an effect of incremental welfare benefits on fertility behavior? A look at the family cap. *39*(2), 295-325.

Kim, Y. I. A. (2014). Lifetime impact of cash transfer on fertility. *Canadian Studies in Population, 41*(1-2), 97-110. doi:10.25336/p64s52

Lalive, R., & Zweimuller, J. (2009). HOW DOES PARENTAL LEAVE AFFECT FERTILITY AND RETURN TO WORK? EVIDENCE FROM TWO NATURAL EXPERIMENTS. *Quarterly Journal of Economics, 124*(3), 1363-1402. doi:10.1162/qjec.2009.124.3.1363

Leibowitz, A. (1990). The response of births to changes in health care costs. *Journal of Human Resources*, 697-711.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., . . . Moher, D. J. P. m. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *6*(7), e1000100.

Lichter, D. T., & Jayakody, R. (2002). Welfare reform: How do we measure success? *Annual Review of Sociology, 28*, 117-141. doi:http://dx.doi.org/10.1146/annurev.soc.28.110601.140845

Liu, Q., & Skans, O. N. (2010). The Duration of Paid Parental Leave and Children's Scholastic Performance. *B E Journal of Economic Analysis & Policy, 10*(1). Retrieved from <Go to ISI>://WOS:000274196700005

Lopoo, L. M., & Raissian, K. M. (2012). Natalist Policies in the United States. *Journal of Policy Analysis and Management, 31*(4), 905-946. doi:10.1002/pam.21646

Machado, M. P., & Sanz-de-Galdeano, A. (2015). Coverage of infertility treatment and fertility outcomes. *Series-Journal of the Spanish Economic Association, 6*(4), 407-439. doi:10.1007/s13209-015-0135-0

McDonald, P. (2006). Low Fertility and the State: The Efficacy of Policy. *Population and Development Review, 32*(3), 485-510. Retrieved from http://www.jstor.org/stable/20058901

Milligan, K. (2005). Subsidizing the stork: New evidence on tax incentives and fertility. *Review of Economics and Statistics, 87*(3), 539-555. doi:10.1162/0034653054638382

Mills, M., Rindfuss, R. R., McDonald, P., te Velde, E., Reproduction, o. b. o. t. E., & Force, S. T. (2011). Why do people postpone parenthood? Reasons and social policy incentives. *Human Reproduction Update, 17*(6), 848-860. doi:10.1093/humupd/dmr026 %J Human Reproduction Update

Moher D, L. A., Tetzlaff J, Altman DG, The PRISMA Group. (2009). Prereferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Plos Medicine, 6*(7), e1000097. doi:https://doi.org/10.1371/journal.pmed.1000097

Neyer, G., & Andersson, G. (2008). Consequences of Family Policies on Childbearing Behavior: Effects or Artifacts? *Population and Development Review, 34*(4), 699-724. Retrieved from http://www.jstor.org/stable/25434736

Olivetti, C., & Petrongolo, B. (2017). The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries. *Journal of Economic Perspectives, 31*(1), 205-230. doi:10.1257/jep.31.1.205

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. J. S. r. (2016). Rayyan—a web and mobile app for systematic reviews. *5*(1), 210.

Parent, D., & Wang, L. (2007). Tax incentives and fertility in Canada: quantum vs tempo effects. *Canadian Journal of Economics-Revue Canadienne D Economique, 40*(2), 371-400. doi:10.1111/j.1365-2966.2007.00413.x

Pirog, M. A., & Ziol-Guest, K. M. (2006). Child Support Enforcement: Programs and Policies, Impacts and Questions. *Journal of Policy Analysis and Management, 25*(4), 943-990. Retrieved from http://www.jstor.org/stable/30162769

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M.. & Duffy, S. . (2006). *Guidance on the conduct of narrative synthesis in systematic reviews.*

Raute, A. (2019). Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits. *Journal of Public Economics, 169*, 203-222. doi:10.1016/j.jpubeco.2018.07.010

Riphahn, R. T., & Wiynck, F. (2017). Fertility effects of child benefits. *Journal of Population Economics, 30*(4), 1135-1184. doi:10.1007/s00148-017-0647-y

Robins, P. K., & Fronstin, P. (1996). Welfare benefits and birth decisions of never-married women. *Population Research and Policy Review, 15*(1), 21-43.

Ryan, R., & CCRG. (2016). Cochrane Consumers and Communication Review Group: data synthesis and analysis. . http://cccrg.cochrane.org, June 2016

Rønsen, M., & Skrede, K. J. D. r. (2010). Can public policies sustain fertility in the Nordic countries?: Lessons from the past and questions for the future. *22*, 321-346.

Sabia, J. J. (2008). Blacks and the family cap: pregnancy, abortion, and spillovers. *Journal of Population Economics, 21*(1), 111-134. doi:10.1007/s00148-005-0049-4

Schmidt, L. (2005). Infertility insurance mandates and fertility. *American Economic Review, 95*(2), 204-208.

Schmidt, L. (2007). Effects of infertility insurance mandates on fertility. *Journal of Health Economics, 26*(3), 431-446. doi:10.1016/j.jhealeco.2006.10.012

Schultz, S. J. D. S. J. o. S. T. (2015). Reproducing the nation: the new German population policy and the concept of demographization. *16*(3), 337-361.

Tach, L., & Edin, K. (2017). The Social Safety Net After Welfare Reform: Recent Developments and Consequences for Household Dynamics. In K. S. Cook & D. S. Massey (Eds.), *Annual Review of Sociology, Vol 43* (Vol. 43, pp. 541-561).

Thévenon, O., & Gauthier, A. H. (2011). Family policies in developed countries: a 'fertility-booster' with side-effects. *Community, Work & Family, 14*(2), 197-216. doi:10.1080/13668803.2011.571400

Thévenon, O., & Luci, A. (2012). Reconciling Work, Family and Child Outcomes: What Implications for Family Support Policies? *Population Research and Policy Review, 31*(6), 855-882. doi:http://dx.doi.org/10.1007/s11113-012-9254-5

United Nations, P. D. (2018). *World Population Policies 2015*. Retrieved from New York:

Welteke, C., & Wrohlich, K. (2019). Peer effects in parental leave decisions. *Labour Economics, 57*, 146-163. doi:https://doi.org/10.1016/j.labeco.2019.02.008

Wood, J., Neels, K., & Vergauwen, J. (2016). Economic and Institutional Context and Second Births in Seven European Countries. *Population Research and Policy Review, 35*(3), 305-325. doi:http://dx.doi.org/10.1007/s11113-016-9389-x

Zavodny, M., & Bitler, M. P. (2010). The effect of Medicaid eligibility expansions on fertility. *Social Science & Medicine, 71*(5), 918-924. doi:10.1016/j.socscimed.2010.05.046

Figures and Tables

Table 1: PICOS for inclusion and exclusion.

CRITERIA	INCLUSION	EXCLUSION
PARTICIPANTS (POPULATION)	1. Populations of nations fully located in Europe (excluding e.g. Turkey and Russia), Northern America (Canada and the US) and Australia. 2. Women or men of childbearing age during the intervention.	1. Teenage pregnancies. 2. Romania, due to a particularly coercive pro-natalist regime under Ceucesco that may generally limit external validity.
INTERVENTIONS	1. Intervention is a policy, implemented at the national, regional or local level. 2. Intervention happened after 1970. 3. The intervention affects the fertility choices of the population.	1. The intervention directly limits participants free choice by restricting access to contraception or abortion. 2. The intervention effects on fertility are unduly complex or indirect, making the intervention an obviously inefficient means of achieving higher fertility.
COMPARATOR/ CONTROL	1. The introduction/revocation of a policy is compared to the absence/presence of the same policy. 2. Modifications of a policy are compared to the same policy in its previous form. 3. Two different policy treatments are compared.	
OUTCOMES	1. Birth rates measured at aggregate (sub-national) level. 2. Birth probabilities measured at individual level. 3. Period ("timing") measures. 4. Cohort ("quantum") measures.	1. Outcome is measured at country level.
STUDY DESIGN	1. Field experiments 2. Quasi-experiments: difference-in-differences, regression discontinuity and instrumental variable design, and any combination of these 3. Two-way fixed effects, or region fixed effects with detailed controls for period and cohort.	1. Observational studies that do not use the strategies mentioned for causal identification 2. Fixed effects are measured at a higher level than treatment.

Note: for further details, see protocol (Bergsvik et al., 2019).

Table 2: Studies of parental leave [PRELIMINARY]

AUTHORS	INTERVENTION	COUNTRY (AFFECTED), IMPL.	MAIN (SECONDARY) OUTCOME	STRATIFICATION	METHOD & RESULTS
CANNONIER (2014)	Introduction of 12 weeks unpaid job-protected leave through Family and Medical Leave Act	USA 1993	Birth probability eligible vs. ineligible women (until 2010)	Sector; Race and Ethnicity; Education	DiD; Timing effect; Earlier births
DAHL, LØKEN, MOGSTAD, SALVANES (2016)	Six Parental Leave extensions (total increase 17 weeks, from 18 to 35)	Norway 1987-1992	Several; Number of children born to a mother 14 years after reform	-	RD Small effect only in 1992; No general effect
LIU, SKANS (2010)	Parental Leave extension (12 to 15 months)	Sweden 1988/89	Children's school performance at age 16 (Timing and number of future siblings + several)	Education	DiD; No general effect; Small increase in prob.of another child within 18 months; Driven by highly educ. mothers
LALIVE, ZWEIMULLER (2009)	Two Parental Leave reforms (flat rate benefit). 1990: 12->24 months + longer speed premium, 1996: 24->18 months + shorter speed premium	Austria 1990 & 1996	Higher order births in short run (3 years) and long run (10 years)	Income; Occupation	RD; Positive short run and long run effects
COOLS, FIVA, KIRKEBØEN (2015)	Introduction of 4-week father's quota (compared to 4-week expansion without reserving share for father)	Norway 1993	Several; Parent's number of children 14 years after reform and spacing	Education	DiD; No effect on fertility
HART, ANDERSEN, DRANGE (2019)*	Extension of father's quota from 6 to 10 weeks	Norway 2009	Subsequent fertility and union stability (earnings)	Parity; Education; Age;	RD; No effect [more details?]
FARRE, GONZALEZ (2018)*	Introduction of 2 weeks paid paternity leave	Spain 2007	Birth spacing and probability of another child within 6 years	Age	RD, DiD; Longer spacing; Neg. effects, driven by mothers > 30
CYGAN-REHM (2016)	Maternity Leave from means tested to earnings related (+ grace period changes)	Germany (West) 2007	Higher order births within 12/21/24/33/36/45/48/57 months	Employment; Eligible for old benefit; Earnings	DiD; Timing effects in line with reform's heterogeneous economic incentives; Neg. persisting effects if low-income; Weak temporary effects if reform winner

RAUTE (2019)	Maternity Leave from means tested to earnings related (+ grace period changes)	Germany 2007	Births high earning vs low earning women; Intensive & extensive margin (5 years)	Age; Parity	DiD; Highly educated more likely to have 1st and 2nd child

ANG MISSING

*Working papers

Table 3: Studies of childcare [PRELIMINARY]

AUTHORS	INTERVENTION	COUNTRY (AFFECTED), IMPL.	MAIN (SECONDARY) OUTCOME	STRATIFICATION	METHOD & RESULTS
RINDFUSS, GUILKEY, MORGAN, KRAVDAL, GUZZO (2007)	Increase in % pre-school-age children in childcare centers	Norway 1973-1998	First birth timing	Age	Two-way fixed effects; Earlier transition to motherhood, and higher probability of becoming mother at every age
RINDFUSS, GUILKEY, MORGAN, KRAVDAL (2010)	Increase in % pre-school-age children in childcare centers	Norway 1973-1998	Total number of children born until age 35	Parity	Two-way fixed effects; Positive effect for all parities, strongest for 2nd and 3rd births
WOOD & NEELS (2019)	Increase in childcare places for 0-3 year olds	Belgium 2002-2005	Probability of having a child among dual-earner couples	Parity	Municipality fixed effects; Positive effect on birth hazard for all parities
BAUERNSCHUSTER, HENER, RAINER (2016)	Expansion of public childcare slots for children under age 3	Germany 2005-2008 (West)	Births per 1000 women aged 15-44; Age-specific birth rates (health outcome of newborn)	Age; Marital status; Parity (among the married)	Generalized DiD; Positive effect on birth rates, driven by married, strongest for 2nd and 3rd births
GATHMANN, SASS (2018)	"Home care subsidy" reform increased price of choosing public childcare compared to home care	Germany 2006 (East German state Thuringia)	Childcare choices (having another child)	Parity; Family status; Education; Income; Citizenship	DiD; No general effect for eligible families; Small positive effects if 2+ children; Stronger if single, low-income, foreign; Discourages 1st births
MORK, SJOGREN, SVALERYD (2013)	Childcare reform standardized childcare fees and imposed a price cap, (1998 announcement, 2002 implementation)	Sweden 2002	Child births per 1000 women	Municipality; Household type (children + income); Voting patterns	DiD; Early positive effect on 1st births, particularly if low-income; 2nd births postponed; Higher order births positive price effect, neg. income effect

BATTISTIN, DE NADAI, PADULA (2015)*	Pension reforms delayed retirement = reduces availability of grandparental care	Italy 1992-2001	Fertility of the offspring	Age; Family tie strength;	RD; Negative effects on offspring's fertility; Varies by tie strength; Formal childcare can attenuate effect

*Working papers

Table 4: Studies on health services [PRELIMINARY]

AUTHORS	INTERVENTION	COUNTRY (YEAR); AFFECTED	MAIN (SECONDARY) OUTCOME	STRATIFICATION	METHOD AND RESULTS
ABRAMOWITZ (2018)	ACA insurance: Reduced cost of conception, birth and abortion	USA (2010); Young adults (20-25) with insured parents	Prob. birth in 12 months; (Contr. use; Trying to get preg.; Abortions)	Age; marital status	DiD. Decrease in births; increase in likelihood of trying to get pregnant. No effect on abortions. Indication of effect on long-term contraceptives. Pre-trend plots and tests
APOSTOLOVA-MIHAYLOVA & YELOWITZ (2018)	Greater accessibility to Medicaid	USA (2006); Massachutes treated	Yearly probability of birth	Age; marital status	DiD with individual level data. Pos. Effect among married women aged 20-34:(1%); neg. Effect among unmarried women (8%). Robustness include excluding movers, excluding control states with minor reforms and changing age brackets.
DELEIRE (2011)	Medicaid insurance for perinatal health, pregnancy and child health	USA (1985-96)	Ln(quarterly birth rates)	Race, marital status, education	2WFE, aggregated data. Medicaid measured as simulated fraction available. No robust relationship
GROVES, HAMERSMA & LOPOO (2018)	Medicaid insurance for perinatal health, pregnancy and child health	USA (1987-97)	Ln(quarterly birth rates)	Race, marital status, education, age, parity	2W FE with aggregated data. Pos. Effect on higher order births for HS educ. Women across race Extensive, incl. limitation of sample to federally initiated changes to avoid self selection. Measured as Medicaid treshold rel. to federal poverty limit
HEIM (2018)	ACA insurance: Reduced cost of conception, birth and abortion	USA (2010); Young adults (20-25) with insured parents	Conception res. In live birth;	Parental income; marital status; parity; postsecondary enrollment	DiD with younger (untreated) as control. Modest decrease in fertility (ITT 7-11%). Robustness include tests for differential pre-trends.
JOYCE ET AL.	Medicaid expansions	USA (ca. 1987-91)	Birth rates; (Abortion rates; Abortion ratio (to births))		2W FE. Increase in births concentrated among whites
LEIBOWITZ (1990)	Free medical care vs. cost sharing insurance	USA (1974-79); Families in six US cities	Births during experiment; (Yearly birth p.)		Experiment. 29% increase in births. Strongest increase after 2-3 years
MACHADO (2015)	Infertility treatment insurance	USA (1979-2001)	Completed fertility; (Age at 1st birth)		2W FE/DiD. No effect completed fert., some delay of first birth

SCHMIDT (2005)	Intefertility treatment: State mandate to provide insurance vs. No mandate	USA (1985-99); Fifteen treatment states	Ln(first birth) rate	Age>35; race	DiDiD with aggregated data. 32% increase among women over 35, concentrated among whites
SCHMIDT (2007)	State mandate to provide insurance vs. no mandate. Additionally: Strong or weak mandate, IVF covered or not, health insurance covered proportion of pop.	USA (1981-99); Fifteen treatment states	Ln(first birth) rate; (Ln(higher order birth rate))	Age>35; race	DiDiD with aggregated data. Positive effect > age 35 among whites only. No eff. At higher parities Not dependent on strength of mandate Stronger with large pop. Covered. Robustness include state specific trends and restr. Time series
ZAVODNY (2010)	Medicaid availability: Insurance for perinatal health, pregnancy and child health Measured as simulated fraction available and expansion treshold	USA (1982-96)	Ln(quarterly birth rates); (Abortion rates)	Race, marital status, education (births only)	2W FE with aggregated data. No overall effect of extensions, possible pos. Effect on low educ. White women. (Restrictions of abortion funding decrease ab. & increase births.)

Table 5: Studies on universal child transfers [PRELIMINARY]

AUTHORS	INTERVENTION	COUNTRY (YEAR); AFFECTED	MAIN (SECONDARY) OUTCOME	STRATIFICATION	METHOD & RESULTS
ANG (2015)	Cash transfers and parental leave benefits; incl ANC	Canada (1980 -->); Quebec	Fertility and female LS	Partnered	Parallell DiD. Pos. Effect of transfers on fertility, stronger for first births
AZMAT (2010)	Tax credit and child deductions; partly conditional on maternal employment	Spain (2003)	Fertility and female LS		DiD for secondary outc. RD or first diff for fertility. Plus 5% fertility
GALLOWAY & HART (2015)	Regional cash transfer	Norway (1989-90); Troms	Fertility (Female LS, Educ.)	Marital status	DiD. Trend modelling and placebo tests. Positive effect on first births among unmarried women
GONZALES (2013)	Universal baby bonus (cash)	Spain (2007); 2y residence	Fertility (Abortion; LS; Consumption)		DiD. Pos eff. On fertilty, Temp. lower LS and less purchased childcare. No eff on consumption
KIM (2014)	Allowance for Newborn Children (ANC); age-adjusted exposure to policy	Canada (1988); Quebec	Completed (cohort) fertility	Age	DiD.Main "age adj." measure is endogenous; however robustness w/exog. measure . No effect on completed fertility
MILLIGAN (2005)	Unconditional cash transfer increasing in number of children (ANC)	Canada (1988); Quebec	Fertility rates; Probability of having child; (Cohort fertility)	Parity; family income	DiDiD with lower parities as contr. Trend inspections. Strong positive effect on third births where incentive is strongest
PARENT (2007)	Quebec-specific expansions of family allowance programs	Canada (Mid 1970); Quebec	Completed (cohort) fertility (Children < 6 in hh)	Age; parity	DiDiD with lower parities as contr. Short term pos. Eff; no lasting eff
RIPHAHN (2017)	Reform reduced cost of 1st child for low-earning couples, and increased cost for high-earning couples	Germany (1996); West	Fertility	Education	DiD. High/low ed take turns being treatment and control. Positive effect on 2nd births for high ed; unexpected neg. Effect on 1st births for low ed

Table 6: Studies on welfare[PRELIMINARY]

AUTHORS	INTERVENTION	COUNTRY (YEAR), AFFECTED	MAIN (SECONDARY) OUTCOME	STRATIFICATION	METHOD & RESULTS
BAUGHMAN & DICKERT-CONLIN (2003)	EITC	USA (1990-99)	First birth rate	Race; marital status	2W FE; Small neg. Effect among whites; small pos. Effects among non-whites
BAUGHMAN & DICKERT-CONLIN (2003)	EITC	USA (1990s)	Birth rates	Race; educ.; parity; marital status	2W FE; Small negative effects for higher-order whites. Tests with state trends and lags.
BREWER, RATCLIFFE & SMITH (2012)	UK Welfare-to-work	UK (1999); low educ. Hh.	Fertility (contraception)	Single and coupled; parity	DiD; Increase in births among coupled women with lower ed. (higher ed. Comparison). Reduction in contraception among low ed.
CAMASSO & JAGANNATHAN (2009)	Family cap, cross state comparison	USA (1980-2000); unmarried	Nonmartial birth rate; (abortion rate; illegitimacy ratio; nonmarital preg. rate)	Race; medicaid funds abortions	FE. Reduction if Medicaid abortions are available; concentrated in states with many blacks. Mediated by abortions. State trends and controls for other welfare changes.
CAMASSO & JAGANNATHAN (2016)	Family cap, cross state comparison	USA (1980-2010); unmarried	Nonmartial birth rate; (abortion rate; illegitimacy ratio; nonmarital preg. rate)	Race; medicaid funds abortions	FE. Reduction if Medicaid abortions are available; concentrated in states with many blacks. Mediated by pregnancies. State trends and controls for other welfare changes.
CAMASSO (2004)	NJ Family cap: Nets out job training eff.	USA (1992-97); mothers on welfare in NJ	Birth of new child (abortion, contraception, sterilization)	New or old in welf. Syst; race	Experiment. Neg. Eff of famcap for short term recipients Long term eff of JOBs training
DYER & FAIRLIE (2004)	Family cap, cross state comparison	USA (1990s); less educated single moms	Nonmarital births		DiD. No robust effect. Triple diff. With marital fertility as control. Robustness includes trend modelling + trend inspection and contr. group sensitivity.
FEIN (2001)	ABC program -- includes job training and parent training	USA (1995-96); Delaware	Marriage; Fertility; (exp. Of marriage and fertility)	Age; parity; marital status; schooling; years of welfare	Experiment. No effect on fertility; effect on fertility plans
FRANCESCONI (2006)	UK	UK (1999); single mothers	LF particip; (paid childcare use; marriage; fertilty)	Parity	DiD. Insignificant neg. effect on fertility. Trends modelled; single women or single w. without ed are control group

GROGGER & BRONARS (2001)	Other welfare	USA (1980-???); unwed mothers	Next birth; marriage	Race	Twinning. White mothers postpone marriage; black mothers have child sooner. Only current, not marginal, benefits influence fert. Relative imp of tw. If benefits are high vs.Low
GROVES (2018)	Subsidies to students; conditional on not married	USA (1982-85)	Ever/age at married/divorced/child	Gender	DiD. Paternal death treatment proxy. Post-reform controls after phase out. Test of balance on covariates. No effect on ever children, Increases age at 1st birth
HOFFMAN & FOSTER (2000)	AFDC	USA (1980s-1991(?)); unmarried women <23 y	Fertility	Marital status; age	2W FE utilizing variation in AFDC across time and states. Pos eff. Among women in early 20s nonmarital. Depends strongly on specification of FE!
HOFMANN & HOHMEYER (2013)	Unemployment benefit reform	Germany (2003)	Pregnancy probability	Breadw. mod.; hh inc, parity	IV. Uses the announcement of less generous benefits as instrument for ec. Uncertainty. Perf. Collin. With period, but placebos look good. Strong ec. Worries reduce pregnancy prob.
HORVATH-ROSE & PETERS (2008)	Family cap, cross state comparison	USA (1984-99)	Nonmarital birth rate; (Marital birth rate)	Race; age; marital status	FE. Negative effects on nonmarital; Pos eff. On marital indicates endogeneity. State trends and controls for other changes.
JAGANNATHAN & CAMASSO (2003)	NJ Family cap -- package effect	USA (1992-97); mothers on welfare in NJ	Birth of new child	Race; racial density; New or old in welf. Syst	Experiment. Negative effect for black women who are new in welfare system and live in non-black areas
JAGANNATHAN (2010)	NJ Family cap -- instruments effect of cash transfer change	USA (1992-97); mothers on welfare in NJ	Birth of new child	New or old in welf. Syst	Experiment combined with IV to test per dollar effect of family cap. Weak negative effect for black women who are new in welfare syst. Mon. Eff. Explains 2.5% of total reform effect
JAGANNATHAN, CAMASSO & KILLINGSWORTH (2004)	NJ Family cap -- package effect	USA (1992-97); mothers on welfare in NJ	Birth of new child	Race; racial density New or old in welf. Syst	Experiment. Negative effect for black women who are new in welfare system and live in non-black areas
JOYCE, KAESTNER, KORENMAN, HENSHAW (2004)	Family cap, cross state comparison	USA (1992-99); higher order births, low educ.	Ln(Higher-order births): For teens: all For women 20-34: Nonmarital; (Abortion rates)	Race	DiDiD using first births as control. Higher order births decline relatively, but not more when family caps are implemented. Inspection of pre-trends

KEARNEY (2004)	Family cap, cross state comparison	USA (1989-98); higher order births	Birth rate	Marital status; Education; Age Parity	DiD. No significant effect with state trends. Test of pre-trends and correlation between fertility level and implementation.
ROBINS & FROSTIN (1996)	AFDC	USA (1980-88)	Fertility	Educ level; race	2W FE. Pos eff. For <HS, black & Hispanic
SABIA (2008)	Family cap, cross state comparison	(1984-98)	ln(Nonmarital birth rate); (Pregnancy; abortion)	Race	2W FE. Negative effects on black nonmarital fertility; Mechanism is fewer pregnancies, not more abortions. Control for other welfare changes.

Figure 1: PRISMA diagram [PRELIMINARY]