The effects of motherhood*

Simen Markussen[†] Marte Strøm[‡]

Abstract

We use miscarriage as a biological shock to fertility in order to estimate the causal impact of motherhood on labor market outcomes. The number of instruments is increased by exploiting the response-heterogeneity to miscarriage along three dimensions: *time*, *age*, and *birth order*. This allows us to separately identify the effect of the first, second and third child as well as the effects of pregnancy and caretaking for small children. We find each child reduces female earnings by around 18%, only part of it due to reduced work hours. We find no evidence of an adverse health effect of having children.

JEL Classification: C26, I10, J13, J22

*We want to thank Sara Cools, Simen Gaure, Jo Thori Lind, Kalle Moene, Magne Mogstad, Ole Røgeberg, Ragnhild Schreiner and Kjetil Storesletten for their help, advice and suggestions. This paper is a part of the projects "Absenteeism - Disability, Norms and Interventions" (grant 201416), "Disentangling absence patterns" (grant 227103) and "Understanding the gender gap in sickness absence" (grant 218378), financed by the Norwegian Research Council. Data made available by Statistics Norway have been essential for the research project. It is also part of the research activities at the Center of Equality, Social Organization, and Performance (ESOP) at the Department of Economics at the University of Oslo. ESOP is supported by the Research Council of Norway.

[†]Ragnar Frisch Centre for Economic Research. E-mail: simen.markussen@frisch.uio.no

[‡]Institute for Social Research and ESOP, University of Oslo. E-mail: marte.strom@samfunnsforskning.no

1 Introduction

The relationship between fertility and female labor market outcomes is a longstanding theoretical and empirical problem in economics. Understanding the relationship is important for several reasons. First, if having fewer children causes an increase in labor-force participation, reduced fertility can potentially explain a large share of the increase in female labor force participation in western societies (Goldin (1990)). Second, if having children causes career-interruptions and periods of reduced working hours, it can explain the persistent gender differences in career and wages (Mincer and Polachek, 1974; Korenman and Neumark, 1992; Budig and England, 2001; Harkness and Waldfogel, 2003). Third, if combining work and family obligations lead to health problems for women due to a ”double burden”, having children can partly explain the increasing gender gap in sickness absence that is observed in Scandinavia as in many other countries with high female labor force participation (Åkerlind et al., 1996; Mastekaasa, 2000; Bratberg et al., 2002). The strong theoretical and empirical association between family formation and female labor market outcomes motivates a closer study of the causal relationship underlying these correlations.

Increasing female labor supply is high on the political agenda in most OECD countries and public resources are increasingly targeted at policies facilitating the combination of work and family, like paid parental leave and subsidized kindergartens. In the Scandinavian countries, these welfare policies are extensively used. The policies are arguably successful: Scandinavia have among the highest rates of female labor force participation and at the same time have fertility rates well above the OECD average. However, Scandinavian women work part-time to a large degree and have high sickness absence rates compared to men. This paper aims to shed light on the role of motherhood in explaining these patterns.

An empirical investigation of the causal relationship between motherhood and labor market outcomes is complicated by the endogeneity of fertility decisions. In the existing research literature only a few sources of exogenous variation in fertility outcomes have been exploited. The relationship between number of children and female labor supply is studied by Rosenzweig and Wolpin (1980) and Angrist and Evans (1998), who are the first to use twin births and the sex composition of the first two children respectively to instrument the effect of the second and the third child, respectively, on female labor supply. They (and several later papers using the same strategy) find that an additional child has a significant impact on female labor supply and earnings. The results are however not directly transferable to the the first child.

This paper contributes to the literature by estimating the marginal effect of the first three children, as well as heterogeneity with respect to the age of the child. Our results contribute to a more comprehensive understanding of the role of motherhood; the quantitative importance of having the first child compared to the second and the third, and whether the effects are long term or temporarily related to the age of the child. We use miscarriage as a biological shock to fertility in order to estimate the causal impact of motherhood on female labor market outcomes the first five years after birth. A miscarriage, compared to a birth, leads to an altered fertility pattern in the years following in the sense that the timing of pregnancies and births are different between the two. They are simply ”out of sync” with each other, a property we use for identification. The ”non-treated” women who get a child the first period continue having more children and have older children than the ”treated” who miscarried in the first period and therefore got the first child as well as the next children later. To identify a meaningful effect of having e.g. the first child, we therefore need to correct for the fact that both the treated and the non-treated are again pregnant or have more children at different ages. The six included variables defining motherhood is the separate effect of having the first, second and the third child, the effect of pregnancy, having an infant and having a toddler (1-3 years old). We use detailed, high-quality administrative data covering the whole Norwegian population and use all registered births and miscarriages in the years 2001-2004. We measure outcomes up to five years after planned birth (from 2001-2009).

With six endogenous fertility variables that are all affected by a miscarriage, we need at least six instrumental variables to identify the causal effects. With few other sources of exogenous variation in fertility outcomes, we increase the number of instruments by exploiting the heterogeneity of treatment response to miscarriage across three dimensions: *time*, *age*, and *birth order*. Firstly, we estimate a separate first stage for all the five years following planned birth. The first stage impact of a miscarriage on fertility outcomes are different in each year, because the group that miscarries has an increasing probability of having a child. This means that we obtain five observations of the first stage relationship between miscarriage and each fertility outcome. The same can be done for different birth numbers; the first-stage effect of a miscarriage on the probability of having a child in a subsequent period is different if it happens before the first, the second or the third child, and it also affects different margins. Lastly, we use the fact that there is also differences in the first-stage across age at planned birth. Older mothers have a smaller biological probability of getting pregnant than mothers at the peak of their fertile age, which means that the first-stage effect of a miscarriage on the probability of having a child the next periods will vary across age-groups.

Combined, the heterogeneity in the first stage treatment response can be treated as multiple instruments (”quasi-experiments”) to estimate a local average treatment effect of each fertility outcome. In practice, the model is estimated using the interaction of miscarriage and group indicators (group being all unique combinations of time period after planned birth, age at first birth, and birth number) as instruments, including group indicators as control variables in the second stage, to ensure that we only compare differences over the fertility shock (miscarriage) within each group. Selection into miscarriage that depends on group variables are thereby accounted for.

The method has similarities to Angrist and Krueger (1991) and their quarter of birth-instrument for schooling. They increase the number of instruments by interacting quarter of birth with year and state indicators in order to increase the efficiency of the 2SLS estimator. They thereby use the variation in quarter of birth within one year instead of across all years. Quarter of birth may have a different impact depending on which year the child is born - and year-specific quarter of birth instruments can therefore provide more variation that increases the efficiency of the estimator. We use the increased number of instruments for a different purpose than increasing the efficiency of the estimator; it enables us to identify the effect of more than one causal variable affected by the initial instrument. When a random event has an effect on several causal variables at the same time, all of them have to be included in the analysis to get unbiased estimates. Our approach is useful when there are no (few) available additional instruments, but there is sufficient heterogeneity in the first stage response across groups that these can be treated as independent observations of the relationship between the instrument and the causal variables of interest. To test linear independence of the instruments, we propose to use a test of near collinearity suggested by Belsley et al. (1980).

The paper also contributes to the small, but increasing, literature using biological shocks to estimate the effect of fertility variables. Miscarriage has previously been used as an instrumental variable for the timing of birth, measuring the long-term effect of having children later; Hotz et al. (1997, 2005) study the effect of teenage childbearing, Miller (2011) study the effect of having children later in the career for life-time earnings and Buckles and Munnich (2012) study the effect of spacing of siblings for child school performance (identified by miscarriages before the second child)[1]. A concern in this literature, is that miscarriage in itself has a negative own effect on the outcomes of interest. An additional advantage of our approach is that we are able to account for such a direct effect. Because we use interaction terms as instruments, we are able to include the original basic instrument miscarriage among the covariates, thereby controlling for a direct (common) effect of the miscarriage (Kolesár et al., 2011). A similar instrument to miscarriage, is to use infertility measures to estimate the effect of having children. Infertility has been used by among others Cristia (2008); Agüero and Marks (2008, 2011). Cristia (2008) finds large effects on female labor supply of having the first child on US data, while Agüero and Marks (2008, 2011) find no association using data from Latin-American countries (Agüero and Marks, 2008) and 26 developing countries (Agüero and Marks, 2011).

We find that each of the first children has a large impact on female earnings, which on average are reduced by 18%. The largest decrease in earnings are driven by women with average earnings reducing their labor supply and thereby earn less. There are small effects on

[1]Previous literature using miscarriage as an instrument for the timing of children do not fully take into account the effect of miscarriage on e.g. completed fertility, and the estimated timing-effect can therefore partly include an effect of total number of children. In addition, this literature only include those who actually end up having a child. This can be considered an endogenous conditioning of the sample as a significant share that experience a miscarriage end up having no children.

earnings for those who are marginally employed or women in the higher end of the earnings distribution. Labor force participation is not affected much by having children, indicating that in the Norwegian context, the combination of market work and family obligations is clearly feasible. The intensity of market work is however reduced. Women work on average 2 hours less per week per child, and the effect does not decrease (much) when the child grows older. Motherhood therefore plays a significant role in explaining female part-time work. We find no evidence of an adverse health effect of having children as neither sickness absence nor disability seem to increase due to motherhood.

2 Modelling strategy

The aim of this paper is to estimate the causal effect of motherhood C_i on labor market outcomes Y_i as specified by (1).

$$\mathrm{Y}_i = \alpha + \mathrm{C}_i\beta + X_i b + u_i \tag{1}$$

The causal relationship from motherhood to labor market outcomes cannot be estimated directly using OLS or similar methods. First, we have a problem of time-invariant selection into motherhood. Women who become mothers can be of a different type than women who do not become mothers, and the type may influence labor market outcomes. Selection on such permanent characteristics can be solved by the use of panel data and within-person estimation. There might however also be time-varying unobserved characteristics of the individual, such as job-motivation and health, affecting both fertility outcomes and labor market outcomes. The most common strategy for solving this problem is the use of instrumental variables, and this is the strategy pursued also in this paper.

Before presenting our model, we need to specify *motherhood*. For most purposes motherhood is defined as having children. For estimation purposes however, several aspects of motherhood may turn out important and need to be specified in greater detail. First of all, the number of children may be of great importance for female labour supply. Also institutional arrangements, such as maternal leave and access to child care, makes it relevant to not only pay respect to the mere existence of children but also their age. Finally, being pregnant is also an important fertility outcome (to be) and may have large impact on labor market outcomes via health as well as employment incentives in order to qualify for paid maternity leave[2].

We model motherhood to consist of a long-term part, the difference between having and not having children, and a short-term part, depending on the age of the child(ren). In our main specification motherhood C_i then consists of six fertility outcomes: (i) The effect of having the first, (ii) second and (iii) third child, (iv) the effect of being pregnant, (v) of having an infant (a child less than one year old) and (vi) having a toddler (a child between 1 and 3 years old).

We use a biological shock to fertility, miscarriage, to identify the causal effect of each of these six dimensions of motherhood on labor market outcomes. We do this by exploiting the highly heterogeneous response to this shock along three dimensions: time, age and birth order, presented in turn below.

Having a miscarriage, compared to a birth, causes a non-synchronicity in fertility outcomes in the succeeding periods, illustrated in Figure 1. Panel (a) shows that the majority that miscarry have children at a later point in time. The share is increasing with time period, and five years after almost 90 % have given birth to a new child. Panel (b) displays the fraction of those who give birth (hollow circles) and those who miscarry (filled circles) that are pregnant again, by year after the planned birth. One year after miscarrying more than 30% of those who miscarried are pregnant, whereas the same number is below 5% for those who gave birth. The later years, the fraction pregnant are quite similar. Panel (c) displays the fraction having a toddler, aged 1-3 years. Those who gave birth at time = 0 are, by definition, having a toddler in years 2 and 3. In year 5 the two groups seem rather similar, but they diverge substantially in years 1 to 4. Combined, this non-synchronocity provides information that can be used to separate the various dimensions of motherhood and identify their impacts on labor market outcomes. The necessary assumption is that at each point in time, the differences in fertility outcomes between those who miscarry in year 0 and those who do not (conditional on X), are caused by the miscarriage.

[2]To have the right to wage-compensated maternity leave, the woman has to work 6 out of the last 10 months before birth.

Figure 1: The impact of miscarriage over time

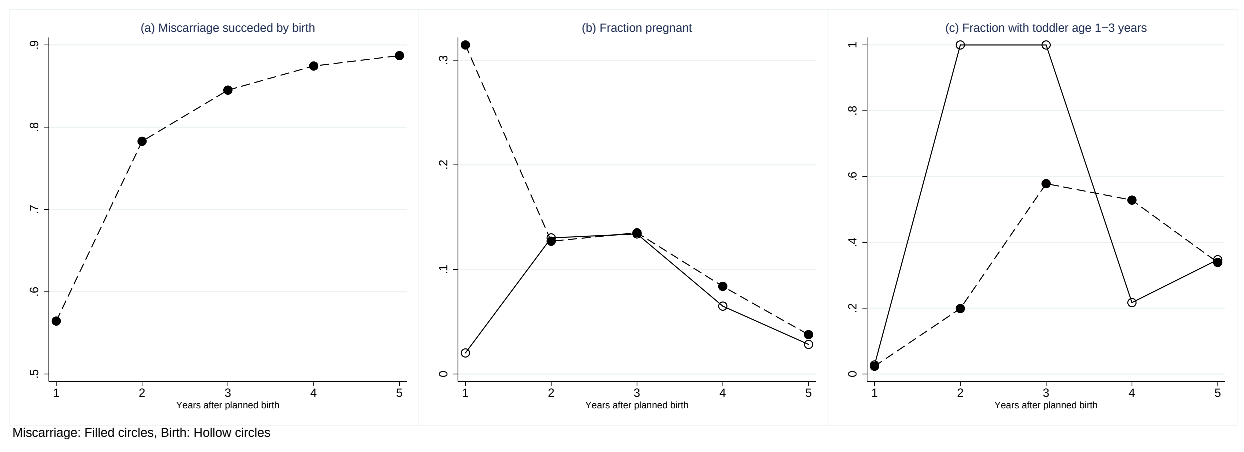
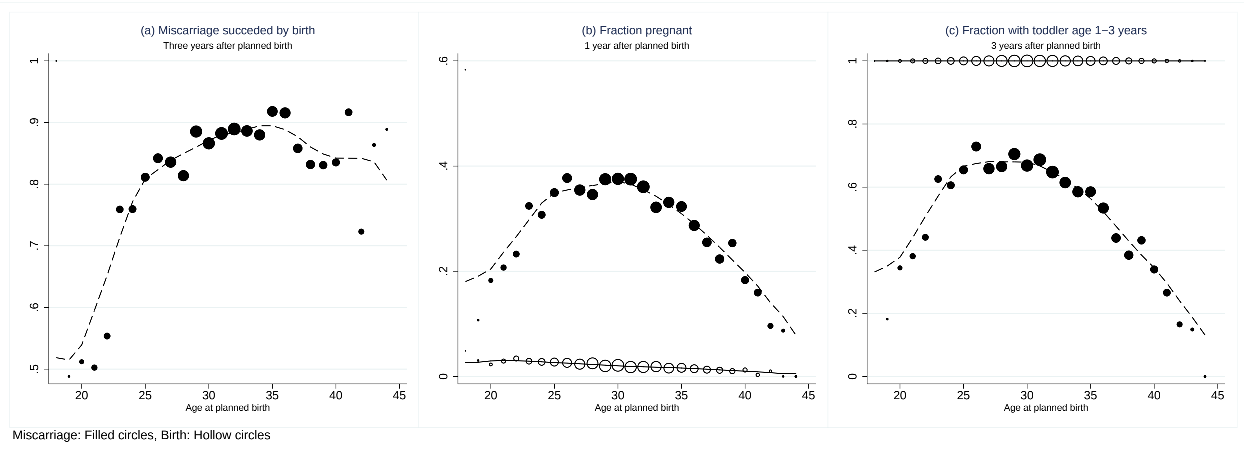


Figure 2: The impact of miscarriage over age



Notes: The size of the circles corresponds to the sample size.

The succeeding fertility outcomes for those who miscarry, compared to those giving birth, also differs substantially with age. This is illustrated in Figure 2. Panel (a) draws the fraction of those miscarrying having a child 3 years after planned birth. There is a hump-shaped pattern where the youngest and the oldest women are least likely to have children after a miscarriage. For the oldest women this probably reflects the, by biology, reduced probability of becoming pregnant when women approach the menopause. For the youngest, it probably reflects that pregnancies in the early twenties are not always planned nor desired. Again we assume that differences in fertility outcomes between those who miscarry and those who give birth, conditional on X are caused by the miscarriage.

Lastly, we treat miscarriage occurring before the first, second and third (planned) child birth as separate events. We can thus identify the marginal effect of each child. The short-term dimensions of motherhood, being pregnant, having an infant and a toddler, are then assumed to have the same effect for each birth-number[3].

Combined, using years since planned birth × age at planned birth × birth order provides us with 5 × 25 × 3 = 375 groups[4]. Formally, our instrumental variable is then the interaction of miscarriage, denoted by z_i and these groups g_{it}. In the case where motherhood is modelled as having K dimensions, we are thus going to estimate $g_{it} \times K$ first-stage coefficients. It is convenient to gather these in K *treatment response vectors* denoted by $\Phi_k = \begin{bmatrix} \phi_{1k} \\ \vdots \\ \phi_{Gk} \end{bmatrix}$. Our empirical model is given by (2) and (3).

$$\mathrm{Y}_{it} = \alpha + g_{it}\Lambda + \theta z_i + \sum_{k=1}^{6} \beta_k \mathrm{C}_{ik} + X_{it}b + u_{it} \tag{2}$$

$$\mathrm{C}_{itk} = \alpha_k + g_{it}\Lambda_k + \Phi_k z_i \times g_{it} + X_{it}b_k + \varepsilon_{itk} \tag{3}$$

where $i = 1,..,N$ denotes persons, $t = 1,..,5$ denotes years after planned birth. The model is estimated using two stage least squares (2SLS) where predicted values from the K first stage equations (3) are inserted for C_{ik} in (2). Standard errors are clustered on the grouping variable g.

In order to provide valid IV-estimates the vector $z_i \times g_{it}$ must be relevant in predicting $C_1 - C_K$ and the exclusion restriction must be satisfied. The question of relevance is an empirical question to which we return thoroughly in the Section 4. As always is the case with IV, the exclusion restriction cannot be formally tested and its validity cannot be proved formally. There are mainly two potential pit-falls regarding the exclusion restriction in our case. The first is related to permanent sorting. We can think of the error term in (2) as consisting of an time-invariant part and time-varying part: $u_{it} = \alpha_i + v_{it}$. In order for the model to provide reli-

[3]Relaxing this assumption and estimating separate age-affects for the different birth-margins, the model loses significance.

[4]There are 27 age groups among those planning their first birth, 25 among those planning their second birth and 23 among those planning their third birth. Potentially we could have 5 × 27 × 3 = 405 groups, but some of the youngest age groups are not represented in the population that is pregnant with their second or third child.

able estimates it must be that α_i is independent of $z_i \times g_{it}$, conditional on the other covariates in (2). Note that the assumption is not that miscarriage is unrelated to unobservable characteristics per se, as z_i is included among the covariates in (2). Instead the necessary assumption is $E[\alpha_i|z_i = 1, g_{it}] - E[\alpha_i|z_i = 0, g_{it}] = E[\alpha_i|z_i = 1] - E[\alpha_i|z_i = 0]$, that there is no unobserved sorting into miscarriage conditional on group g_{it}. For example within an age-group, there are no unobservable characteristics influencing the probability of miscarriage. We attempt to test this assumption by regressing $z_i \times g_{it}$ on realizations of Y_{it} in the past, controlling for g_{it}. This is done in the next section.

Note that the exclusion restriction in our model is weaker than the usual assumption in IV-estimation with one instrumental variable and one causal variable. In the standard IV-model the exclusion restriction is: $E[\alpha_i|z_i = 1] - E[\alpha_i|z_i = 0] = 0$; no correlation between the probability of $z_i = 1$ and unobservable characteristics α_i is allowed. In our model, there can be such correlation or sorting into miscarriage, as long as it is homogeneous across groups g. By including z_i in the model, as a covariate, such sorting is controlled for.

Having a miscarriage can potentially affect women ”directly”, for example through psychological distress, not just through fertility outcomes. In the special case where $E[\alpha_i|z_i] = E[\alpha_i]$, when there is no sorting into miscarriage, we can also identify the direct effect of having a miscarriage on Y_{it}. The parameter identifying this effect is θ in Equation 2. The intuition is that we use the first stage response heterogeneity across groups, net of the mean effect of z_i, as identifying variation. This is different from the usual 2SLS estimation with one instrument, where the mean response to the instrument is the identifying variation. The fact that we can control for a direct (common) effect of the original basic instrument when we use interaction terms as instruments is also shown in Kolesár et al. (2011). In the case with group constant sorting into miscarriage and a direct effect of miscarriage on Y_{it}, the coefficient θ will capture a mix of the two and consequently be hard to interpret.

The second necessary condition for the exclusion restriction to hold is that $z_i \times g_{it}$ is independent of the time-varying part v_{it} of the error term. In other words, conditional on g_{it} and z_{it}, $z_{it} \times g_{it}$ should have no other effect on Y_{it} than through the first stage equations. This assumption is in principle not testable and remains the identifying assumption of this paper.

3 Data and definitions

The starting point of our analysis is the population of all Norwegian women who either gave birth or was registered with a sickness absence spell because of miscarriage in the period from July 2001 and throughout 2004. We need to make a few sample restrictions. First, we only include observations on women who were employed in the period from four weeks before getting pregnant to four months before birth because the data on miscarriage is only available for those employed - and therefore can get sickness leave[5]. Women in our sample therefore have the right to sickness absence from the day they become pregnant to four months before birth (after that, losing the child will not be diagnosed as a miscarriage, but as a *still birth*). Second, we only included women in fertile age (less than 45 years old) and age groups where labor market outcomes are of relevance (at least 18 years old).

The data we use contain start dates and end dates for all sickness absence spells certified by a physician covering the entire population of Norwegians. The Norwegian sickness insurance system covers all employees and the replacement rate during sickness is 100% of yearly earnings, up to a maximum of 83 743 USD or 511 470 NOK (2013 prices). As a main rule, all sickness spells exceeding 3 days must be certified a GP[6]. The data include diagnosis and this is the basis of our measure of miscarriage. The data are unique for our purpose because they do not suffer from bias usual for survey data, like measurement error and selection of who reports an experience of miscarriage[7]. The data on sickness absence spells are merged with linked administrative register data from Statistics Norway providing detailed information on age, education, region of residence, the linking of children to their parents and work status information like being on welfare and participating in the labor force.

[5]An individual has the right to compensated sickness absence if she has worked the last four weeks.

[6]Workers employed by firms which are members of the Agreement for an Inclusive Labor Market, ”IA-avtalen”, can have up to 8 days of self-certified absence days before visiting a physician.

[7]Previous literature using miscarriage as an instrument rely on survey data.

We use two different measures of miscarriage[8]. One is based on diagnosis from the primary care physician (GP)[9], the other is based on diagnosis from the hospital or specialist healthcare[10]. We estimate the model separately for the two different miscarriage types, leaving the GP miscarriages out of the sample when using hospital miscarriages as instrument and vice versa.

Unfortunately for our purpose, pregnancies are not registered, only births and sickness leaves. Hence, we are not able to perfectly observe *when* during a pregnancy a miscarriage took place. It seems however plausible that miscarriages registered with the GP occur early during the pregnancy, without particular physical complications, because that would involve a standard examination at the hospital (and a registered diagnosis from the hospital if treated there). We set the length of pregnancy for this group to 12 weeks at the time of miscarriage and calculate the expected month of birth based on this. The main disadvantage of this measure, is the fact that going to the doctor and obtaining a sickness certificate after a miscarriage is decided by the individual. There is the concern therefore, that those registered with a miscarriage from the GP are a selected group. As we see in Table 1, this group has also on average a longer period of sick leave after the miscarriage than the group with a diagnosis from the hospital.

Our second measure of miscarriage is based on diagnosis from a hospital or a specialist healthcare, which means that the women who miscarry with this diagnosis have been to the hospital because of the miscarriage[11]. The decision to be sent to the hospital is made by the GP due to physical reactions that needs to be examined by a specialist or due to the length of the pregnancy (Hospitals do not receive patients before they have a referral from the GP in Norway). We thus assume that hospital miscarriages occur later in the pregnancy than those leading to a sick leave certified by the GP, and set the length of pregnancy to 17 weeks (the middle between 12 and 22 weeks) at the time of miscarriage.

[8]The international WHO definition of a miscarriage is a pregnancy loss before the 22 week (this is including the two weeks prior to pregnancy - the total length of a pregnancy is 40 weeks). The definitions might differ for different countries, and the practice may also vary between hospitals. The diagnosis codes ICD10 used by hospitals/specialist health care define a miscarriage as between the 6th and the 24th week.

[9]The diagnosis code is ICPC-2, W82.

[10]The diagnosis code is ICD-10, O03.

[11]The diagnosis does not include abortions, which has separate codes in the diagnosis system.

There are pros and cons with hospital-registered miscarriages compared to those registered by the GP. Since later abortions typically involve some kind of medical examination (with or without succeeding treatment) we believe the problem of selection is much less severe for miscarriages registered by the hospital. However, if we expect the unfortunate event of a miscarriage to have long-term effects on those experiencing it, we would believe this problem to be more severe among hospital-registered miscarriages. We return to this is the robustness section below.

Ashcraft and Lang (2006); Ashcraft et al. (2013) point to the role of abortion to the selectivity of the sample. If those who have an abortion is a selected group, the group that miscarries is potentially also a selected group compared to the group that gives birth because some miscarriages are "latent abortions". This concern is larger for the GP miscarriage group that are earlier miscarriages and some of them may therefore be latent abortions. This is not the case for the hospital miscarriages that happen after the time-limit for self-decided abortions (12 weeks in Norway) and thus resemble the group that actually gives birth. For the hospital miscarriages therefore, the effect we estimate is the effect of having compared to not having children for the group that have decided to keep the baby.

The share of women in our data who are registered with a miscarriage is around 5%. This is less than the occurrence of miscarriage referred in the medical literature (around 10% of clinically recognized pregnancies do not result in the delivery of a baby (30% if you count early pregnancy loss) (Wilcox et al., 1988; Wang et al., 2003)). For many, experiencing a miscarriage does not result in a visit to the doctor or sickness absence, and is therefore not registered in our data. The probability of underreporting is highest for early miscarriages that do not include large physical reactions/complications.

Our outcome variables are yearly earnings, weekly hours, sickness absence and other welfare benefits. Yearly earnings include both labor and capital incomes. Weekly hours is a measure of an average working week for the individual's main work relation in the worker-employer registry. In the case of no registered employer, but earnings above certain thresholds, we assume the individual to be self-employed and set hours accordingly. Sickness absence is

Table 1: Observations of birth and miscarriages in our data. All Norwegian women in the period 2001-2004

	Birth	Miscarriage, GP	Miscarriage, hospital
First child			
Observations in data	82,679	1,835	1,824
Employed and aged 18-45	46,681	1,293	1,269
Absence after miscarriage		13.4 days	8.6 days
Postponement of birth		12.2 months	14.1 months
Age	29.2	29.1	30.0
Years of schooling	14.8	14.1	14.8
Second child			
Observations in data	36,036	835	764
Employed and aged 18-45	21,198	637	589
Absence after miscarriage		17.3 days	10.1 days
Postponement of birth		12.4 months	13.1 months
Age	31.9	32.8	33.6
Years of schooling	14.5	14.4	14.7
Third child			
Observations in data	18,792	516	481
Employed and aged 18-45	10,299	365	359
Absence after miscarriage		23.1 days	10.0 days
Postponement of birth		11.6 months	13.5 months
Age	34.0	35.2	35.8
Years of schooling	14.4	13.8	14.4

absences longer than 16 days and is conditional on employment. The measure of other welfare benefits are various social welfare benefits that all citizens are eligible for, regardless of employment. Tables 1 and A.1 shows descriptive statistics for our sample in the year of planned birth.

4 Instrument validity

According to the medical literature, miscarriage is in most cases caused by anomalies in the fetus (e.g chromosomal aberrations) or the mother has a physical defect (uterine anatomic defect) (Kline et al., 1989; Garcia-Enguidanos et al., 2002). The impact of behavioral factors on miscarriage risk are small; extreme behavior like heavy alcohol drinking or drug use can lead to miscarriage, although rare. Among risk factors may also be workplace toxicants (see Garcia-Enguidanos et al. (2002) for a review of risk factors). In addition, miscarriage risk increases with age, particularly in the late thirties. In the estimation, we control for age, nationality, education and two-digits sectoral industry codes (ISIC) at planned birth, controlling in that way for risk-factors correlated with these. Conditional on these covariates, we show that miscarriage is not significantly related to labor market outcomes prior to pregnancy (section 4.1).

Because we do not observe miscarriages per se, but an administrative record of medically-approved sickness absence, there is a potential selectivity into the sample as discussed in the section above. GP registered miscarriages involves more individual choice (because of smaller physical reactions, the individual may decide not to visit the doctor and get sickness absence). This leads to potential underreporting and selection into the sample which is partly confirmed in section 4.1. Hospital registered miscarriages are a priori not likely to be underreported and the sample with hospital miscarriages is thus less likely to be a selected sample. This makes hospital miscarriages our preferred measure of miscarriage.

A particular concern with using miscarriage as instrumental variable is that it may affect women in more ways than through fertility. If she e.g experiences a depression following the miscarriage, this can have negative career consequences. The medical literature compares the period after a miscarriage to a period of grief where feelings of depression/anxiety is most pronounced the first six months after the experience and then wavers off, and is back to ”normal” after a year (Broen et al., 2004; Lok et al., 2010) or six months found by others (Brier, 2008). As shown in Section 2, we control for a direct (common) effect of the miscarriage on the outcomes by including the original instrument as a regressor. We have also calculated frequencies of sickness absence with a diagnosis of depression before birth and in the first years after the miscarriage. The numbers show that the group that miscarry with a diagnosis from the hospital is very similar to the group that gives birth. The group that miscarries with a diagnosis from the GP, however, has a slightly larger overall frequency of psychiatric diagnosis' and also experiences a large jump in the frequency the year after the miscarriage - additional indications of the

GP sample being a more selected sample[12]. Such comparisons are however not an appropriate test for any direct effects of miscarriage on labor supply, since sickness (benefits) only are observed for those working, and having children affect employment. We return to this issue when interpreting results and in the robustness tests, where the model is estimated on a dataset where the first year is excluded.

4.1 Correlation with past outcomes

For the exclusion restriction to be fulfilled, miscarriage must not be correlated to unobservable characteristics conditional on the grouping variable g_{it} and the vector of covariates X_{it}. As a test of this assumption, we estimate a reduced form model on pre-pregnancy labor market outcomes 2 years before birth/miscarriage[13]. The equation is given by (4) where ρ denotes the year of planned birth.

$$Y_{i,\rho-2} = \alpha + g_{i,\rho}\Lambda + \theta z_i + \Psi z_i \times g_{i,\rho} + X_{i,\rho} b + u_{i,\rho-2} \tag{4}$$

Since we use the vector of interactions $z_i \times g_{i,\rho}$ conditional on $g_{i,\rho}$ to identify causal effects, the relevant test is whether the vector Ψ is statistically different from zero prior to the realization of z_i. If the instrumental variable(s) also has predictive power on outcomes prior to the realization of z_i it would indicate unobserved and persistent sorting that would violate the identifying strategy. The results are reported in Table 2. We also report results for the average effect of z_i, θ [14].

Table 2 shows the value of an F-test testing the joint significance of the vector $z_i \times g_{i,t}$, and the p-value shows the level of significance. The results are reassuring. The interactions are far from being significant for any outcome variable in the sample with a miscarriage from the hospital. Notably, for sickness absence and other welfare benefits, there is no association between these health measures and experiencing a miscarriage two years later. The test performs well

[12] Results available upon request

[13] We need to go at least two years back since some of the data are based on calendar years and we want to use a period where we have not conditioned on any outcomes. We also did the same analysis 3 years before planned birth, and the conclusions are largely the same.

[14] We need to assign one value of $z_i \times g_{i,\rho}$ to be the reference category. This is why it is possible also to identify θ.

Table 2: OLS estimates of Equation 4: The reduced form impact of $z_i \times g_{i,\rho}$ on labor market outcomes two years before pregnancy.

	Including individual characteristics				Excluding individual characteristics			
	Earnings	Hours	Absence	Disability	Earnings	Hours	Absence	Disability
Miscarriages registered by hospital/specialist health care								
F-value	0,99	0,81	1,06	1,04	0,99	0,76	1,09	1,06
p-value	0,49	0,86	0,35	0,39	0,51	0,93	0,29	0,35
z - beta	-543,4	0,18	-0,01	0,01	1507,5	0,23	-0,01	0,01
z - se	1991,6	0,94	0,01	0,01	2240,6	0,97	0,01	0,01
z - t-value	-0,27	0,20	-1,16	0,68	0,67	0,23	-1,32	0,39
Miscarriages registered by GP								
F-value	0,85	1,34	1,07	1,22	1,11	1,39	0,98	1,22
p-value	0,80	0,03	0,33	0,10	0,26	0,02	0,51	0,10
z - beta	-2405	0,55	0,02	0,003	-2003,4	0,44	0,03	0,004
z - se	2305,3	1,12	0,01	0,01	2590,7	1,16	0,01	0,01
z - t-value	-1,04	0,49	1,65	0,22	-0,77	0,38	1,82	0,26

Notes: Included covariates are: $g_{i,\rho}$, year at planned birth (4 dummy variables), month at planned birth (12 dummy variables), nationality (5 dummy variables), years of schooling (9 categories) and two-digits sectoral industry codes (ISIC) at planned birth (61 dummy variables).
The F-value refers to an F-test of the significance of the vector Ψ in Equation 4, the P-value shows the level of significance. z-beta refers to the coefficient θ, the mean impact of z_i on the outcome.

even when we do not include covariates in the second part of the table, supporting the assumption that our measures of miscarriage reflects a biological shock, not related or influenced by behavior on the labor market. For the group that miscarries with a diagnosis from the GP, there are signs of a significant relation between miscarriage and weekly hours, both with and without covariates in the regression. This confirms our concern that those who are registered with a miscarriage from the GP are a more selected group. We estimate therefore the model separately for the two different miscarriage groups, and also control for lagged outcome-variables in some specifications.

We also report the coefficient and the standard error for the average effect of z_i on labor market outcomes two years prior to the miscarriage (In the table denoted z - beta). For the hospital sample there is no significant association for the average effect either, which is in line with the existing literature using miscarriage as an instrumental variable. For the GP sample, however, there is a significant relationship between miscarriage and sickness absence in $t-2$. This is further indications that the GP sample are a selected sample that either have poorer health or go more often to the doctor when ill.

4.2 First stage coefficients - heterogeneity in treatment response

For an instrumental variable approach to provide trustworthy results the instruments must be relevant in the sense that they have a substantial impact on the endogenous covariates. In the standard application with one endogenous variable this is tested by a F-test. The rule-of-thumb for "substantial impact" is usually an F-value for the instrument(s) of at least 10 (Stock and Yogo, 2005). With many dummy instruments, there is a particular concern that they are weak and that the coefficients are biased towards the OLS coefficient. This was pointed out in Bound et al. (1995) as a response to the way Angrist and Krueger (1991) increased their number of instruments by interacting quarter of birth with year and state dummies. With a weak first stage, estimates may suffer from finite-sample bias even though the estimating sample is large. With many instruments and more endogenous variables, the F-values need to be larger than the rule of thumb of at least 10. Stock and Yogo (2005) provide critical values to evaluate the result from tests of weak instruments like Cragg-Donald and Kleibergen-Paap, but the maximum number of endogenous variables are 3 in their table. We therefore need to rely on the F-values being ”large” and also look at the individual strength of each instrument (t-values) in addition to Shea's partial R^2.

There does not seem to be a problem of weak instruments when looking at the F-values in Table (3) which are large for all the six endogenous variables. Shea's partial R^2 for the six endogenous variables are not alarmingly small, and comparable to the first stage partial R^2. Each individual t-value is also highly significant[15]. We therefore use all instruments in the main specification to gain efficiency and enough variation to identify the effect of the six different fertility variables. We do however provide results for other combinations of instruments in Section (7).

Strong instruments may not be sufficient in the case of more than one endogenous variable. Consider a hypothetical case where the impact of z_i is equal for all g_i, such that $\phi_{gk} = \phi_k \forall g$. The partial F-test for each of $C_1 - C_K$ may still show that the instrumental variable is strong as long as ϕ is significantly different from zero. Still, in this hypothetical example there is just one instrumental variabel, z_i and the model is not identified.

Formally, what we need is the block matrix of $\mathbf{\Phi} = \left[\theta \ \Phi_1 \ \cdots \ \Phi_K\right]$ to be of full column rank. However, since these are estimated coefficients, bothered with statistical noise, this criterion will in a practical (and numerical) setting nearly always be satisfied. In practice, one may thus estimate (2) and (3) and obtain estimates even if the model is unidentified. Hence, a traditional test for exact collinearity will not serve as a proper test for identification. However, a procedure for detecting *near* collinearity, suggested by Belsley et al. (1980) (BKW) is useful in this respect. Belsley (1991) provides a guide for using their procedure, which is based on a singular value decomposition of the explanatory variables. Using this procedure one obtains guidance on the presence of (and number of) near dependencies in the data. In addition, using a variance decomposition one can see which of the estimated parameters that are involved in the near dependencies to such an extent that one suspect it to be degraded.

[15]Results available upon request.

Table 3: Instrument relevance: F-values and Shea's adjusted partial R^2 from the first stage estimation of Equation 3; the impact of $z_i \times g_{it}$ on the six fertility outcomes.

	F-statistics		Shea's adj. partial R^2	
	Hospital	GP	Hospital	GP
First child	1340,9	1384,7	0,083	0,081
Second child	24,06	24,63	0,008	0,006
Third child	62,07	60,55	0,015	0,013
Pregnancy	16,47	14,02	0,068	0,067
Child 0-1	95,02	77,24	0,015	0,013
Child 1-3	13,61	80,97	0,003	0,001
Highest C.I.	13.61	11.81		

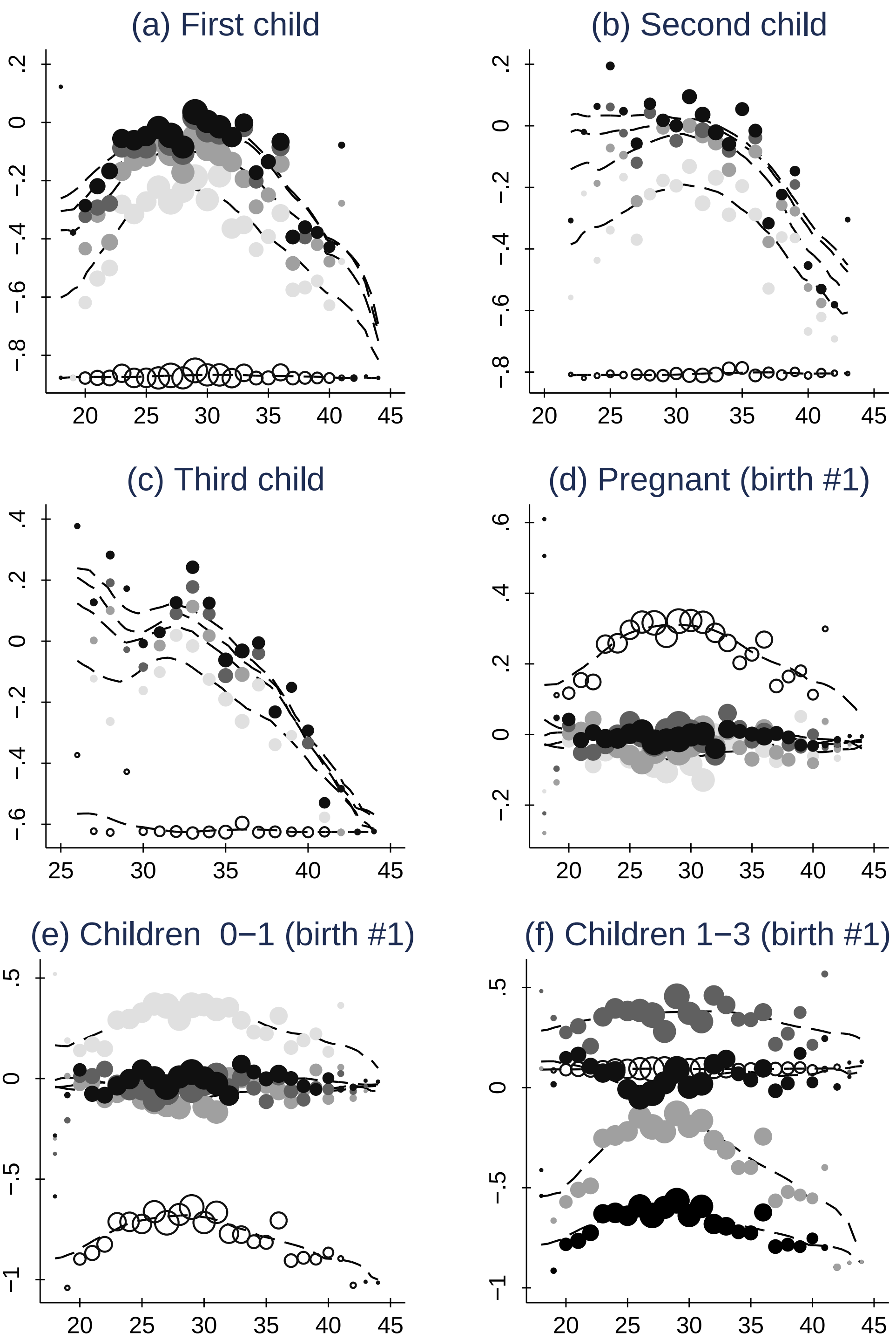
Notes: The table shows F-values from testing the joint significance of the instruments in the first stage regression. Included covariates are: z_i and g_{it}, year at planned birth (4 dummy variables), month at planned birth (12 dummy variables), nationality (5 dummy variables), years of schooling (9 categories) and two-digits sectoral industry codes (ISIC) at planned birth (61 dummy variables).
Condition index numbers (highest C.I.) are results from Belsley et al. (1980) test of near collinearity.

Table (3) displays both the partial F-test values as well as the condition indices from the singular value decomposition suggested by BKW for detecting near collinearities. Note that the singular value decomposition is applied on the estimated first stage coefficients directly $\mathbf{\Phi} = \left[\theta\ \Phi_1\ \cdots\ \Phi_K\right]$. The partial F-tests show that the instrumental variables have sufficient predictive power on each of the endogenous variables. Furthermore, the condition indices are well below 30 which is the critical value described by Belsley et al. (1980), indicating that there are no strong dependencies among the the K vectors of first stage coefficients $\Phi_1 \cdots \Phi_K$.

The basis for identification, the heterogeneous response across g_{it} to the fertility shock z_i is easiest illustrated by plotting the estimated first stage coefficients $\mathbf{\Phi}$ from Equation (3) in a graph, as in Figure (3).

The six different panels show first stage coefficients for the six different endogenous variables in our model; having a first child, having a second child, being pregnant, having an infant and having a toddler. We have one first stage coefficient for each unique combination of years since planned birth $\times$ age at planned birth $\times$ birth order, but for dispositional purposes we have not included all 375 coefficients for each variable. We illustrate years since planned birth by

Figure 3: The first stage impact of $z_i \times g_{it}$ on the six causal variables; (a) The probability of having a first child, (b) The probability of having a second child, (c) The probability of having a third child, (d) The probability of being pregnant, (e) The probability of having an infant, (f) The probability of having a toddler



The figure plots first-stage coefficients Φ_k from Equation 3 using the sample of hospital miscarriages.

color intensity, where white is the first year after planned birth and black is five years after planned birth. Age at planned birth is measured along the x-axis. The size of the dots indicates number of observations underlying the estimate. The dotted lines connect the dots within time-period.

In the first panel (a), we have plotted the first stage coefficients within age-group and time-period for the women who are pregnant with their first child (second and third births have no impact on first births, so they are in any case not relevant here). The reference woman is 30 years old, pregnant with her first child, observed 5 years after the miscarriage. The figure shows how first stage coefficients are heterogenous according to time and age. In the first period, the impact of a miscarriage on the probability of having a first child is -1 (white dots), that is none of those who miscarrry have a child when planned (as is evidently the consequence of a miscarriage). The next period (lightest grey color), the impact of a miscarriage is much smaller, ranging from -.6 to -.2. This means that a large share of those who miscarry one period has a child the next period. In the last period, almost everyone in the agegroups around 30 have their first child, but especially for the oldest age groups, the probability of having a first child is still very small. The heterogeneity with respect to age is very clear.

The second panel (b) shows first stage coefficients for the probability of having a second child for women who are pregnant with their second child. To be able to interpret the coefficients, the reference woman is here a 30 year old, pregnant with her *second* child, observed 5 years after the miscarriage[16]. The same time pattern and heterogeneity in first stage coefficients across age-groups is visible. The probability of having a second child is increasing with time period, and the probabilities are always higher in the 25-35 age groups. Note that the youngest age groups are not represented among those who are pregnant with their second child. We have not plotted the first stage coeffients for first birth mothers in panel (b), but we find small significant first stage effects of having a miscarriage during the pregnancy with the first child on the probability of having a second child. This is because a miscarriage postpones the whole fertility pattern of the mother - it postpones the first child, but also the second child. The marginal effect

[16]This is only for dispositional purposes in this figure, in the etimations we always use a 30 year old, pregnant with her first child, 5 years after the miscarriage as the reference.

of having a second child on labor market outcomes is therefore partly identified by women who miscarry during their first pregnancy.

The third panel (c) shows the first stage coefficients for the probability of having a third child for women who are pregnant with their third child. Also here we change the reference person to a 30 year old, pregnant with her *third* child, observed 5 years after the miscarriage to be able to interpret the coefficients. The time and age pattern is the same also in this group. As for the second child, we did not plot the coefficients for the group that miscarried during their first or second pregnancy, but we do have significant first stage coefficents from these groups also on the probability of having a third child.

The last three panels (d)-(f) shows the first stage coefficients for the proabability of being pregnant, having an infant and having a toddler. The reference woman in all these panels is 30 years old, pregnant with her first child, observed 5 years after the miscarriage. These variables are mainly time varying versions of the three first variables and are included to separate the effect of pregnancy from having no children and having young children from older children. We see from the first stage coefficients that both heterogeneity along the age and the time dimension are important for identifying the effect of these variables.

5 Results

The effects of motherhood on yearly earnings, employment and out of work benefits are displayed in Table 4[17]. Results for the GP sample are reported in table A.2 in the appendix. We find that having children has a large negative impact on mothers' earnings. The marginal effect of each child is almost equally large and reduces yearly earnings by around 10 000 USD (60 000 NOK, 2013 prices) on average. Measured as percentage of earnings in t+1, the reduction in earnings is around 18 percent[18]. Having a toddler reduces earnings even further, around 5 percent. These effects are considerable and have long term consequences e.g directly through a

[17] Out of work benefits includes benefits which are unconditional on employment, i.e.: temporary and permanent disability benefits, unemployment benefits and social assistance/welfare.

[18] Descriptive statistics for the outcome variables in year t+1 and t+5 are in Table A.1 in the appendix.

Table 4: 2SLS estimates of the main model (Equation 2 and 3): The impact of six fertility outcomes on labor market outcomes.

	Unconditional outcomes				Cond. on employment	
	Earnings	Employment	Hours	Benefits	Hours	Sickn. abs.
First child	-10 745***	-0.048*	-3.20***	0.008	-2.20**	-0.01
	(2 813)	(0.027)	(1.08)	(0.022)	(0.10)	(0.02)
Second child	-12 076***	-0.048*	-2.66***	0.010	-1.65*	-0.03*
	(3 224)	(0.029)	(1.14)	(0.024)	(0.11)	(0.02)
Third child	-10 331***	-0.058**	-2.81***	-0.011	-2.09**	-0.03*
	(2 743)	(0.026)	(1.07)	(0.023)	(0.10)	(0.02)
Pregnancy	-6 564	0.012	1.11	-0.066*	2.26	0.36***
	(5 473)	(0.050)	(2.08)	(0.036)	(0.21)	(0.03)
Children 0-1	-3 322***	0.022*	-0.09	-0.039***	-0.19	-0.01
	(999)	(0.012)	(0.49)	(0.010)	(0.03)	(0.01)
Children 1-3	-2 579**	-0.024***	-0.72***	0.011**	0.02	-0.01
	(567)	(0.007)	(0.27)	(0.006)	(0.05)	(0.01)
θ	-2 277*	-0.026***	-0.14	0.001	0.07	-0.01*
	(1 222)	(0.007)	(0.50)	(0.010)	(0.05)	(0.01)
Highest C.I.	13.61	13.61	13.61	13.61	13.5	13.5
J-value, overidentification	315.0	279.8		362.4	283.5	320.0
p-value, overidentification	0.701	0.977		0.099	0.967	0.629
N	401 955	401 955	401 955	401 955	369 909	369 909

Notes: included covariates are: z_i and g_{it}, year at planned birth (4 dummy variables), month at planned birth (12 dummy variables), nationality (5 dummy variables), years of schooling (9 categories) and two-digits sectoral industry codes (ISIC) at planned birth (61 dummy variables). Standard errors clustered on group g are reported in parenthesis. Each column is one regression.
Highest C.I. refers to the Belsley et al. (1980) test for near collinearity.

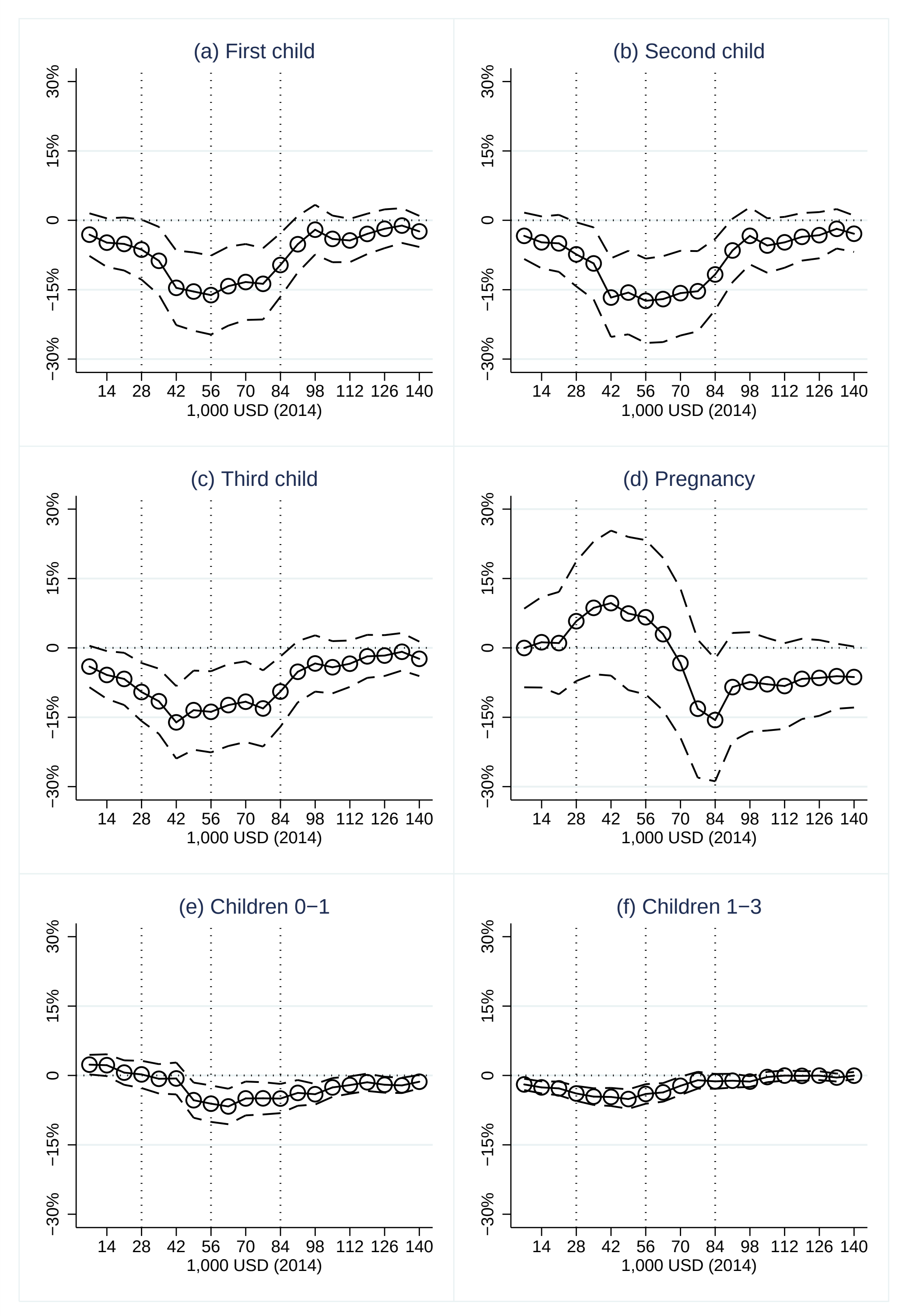
reduction in earned pension rights, or indirectly via the adverse effect of reduced employment on later career opportunities. The drop in earnings when the child is below one year old is probably because many prolong the parental leave period either without pay some months, or by cutting the compensation every month to stretch the period. Pregnancy has no significant effect on earnings.

A part of the fall in earnings is due to a fall in labor force participation. Labor force participation and employment is here defined as having yearly earnings above 1 base amount (BA) in the norwegian social insurance system (in 2013: USD 14 200/ NOK 85 245)[19]. We find that the first three children reduces the employment rate by around 5 percentage points. The marginal effect of each child is almost equally large. The effect on employment is smaller for infants and this is probably because most women in the analysis have the right to paid parental leave - and therefore earn more than 1 BA in the year they are on leave. There is an extra 2.4 percentage points decrease in employment from having a toddler aged between 1 and 3 years.

Figure 4 displays the effect of motherhood on the probability of having earnings above 20 different earnings cut-offs, ranging from USD 7 000 to USD 140 000. Setting the earnings limit low, we see there is no significant effect of having neither the first, second, nor the third child. This means that having children has no significant effect on the probability of having zero earnings. Increasing the earnings limit, the effect of children becomes larger. The effect is biggest around 42 000 USD - 84 000 USD which indicates that the main effect on female labor supply, is the increased probability of working part-time and thereby getting lower earnings. For earnings limits above 84 000 USD, which is approximately the 15th percentile, there is no significant effect of the first three children. Women with high earnings do not reduce their labor supply significantly because of children. The additional effect of having small children is small, but follows the same pattern as the longer term effect of children, that there is an increased probability of having earnings fall below 42 000 USD - 84 000 USD. The effect of pregnancy is not significant, but note that the pattern is consistent with the economic incentives to work during pregnancy to earn the rights to paid parental leave. The maximum limit of

[19]It is not at all clear how to define employment, i.e. how much should one work in order to be defined as employed. Figure 4 relaxes this assumption and displays the effects on 20 different cut-offs.

Figure 4: 2SLS estimates of the main model (Equation 2 and 3; the impact of the six different fertility outcomes on the probability of falling below different earnings thresholds.



Each circle is a point estimate of the impact of falling below the earnings thresholds indicated on the x-axis (each panel (a)-(f) displays the results from 10 different estimations). 95% confidence intervals are indicated by the slashed lines.

compensation is set to 84 000 USD, and we see that for those earning below this limit, earnings (and therefore labor supply) increase during pregnancy.

The total effect on labor supply is captured in the effect on unconditional hours. This is the total drop in hours and includes both the extensive margin (employment) and the intensive margin (weekly hours). The total effect is around three hours, about 9-10 percent. If we condition on employment in column (5), the effect on hours is smaller, but still substantial. Each child reduces female labor supply by on average 2 hours a week, around 6 percent. This is consistent with the high degree of female part-time work in Norway. The effect of each child is almost equally large, so that if a woman has three children, the probability that she works part-time is very high. Because labor supply is here measured by contracted hours in the main work relation, we may underestimate the effect on labor supply if the woman reduces overtime or cuts down on extra jobs. The effects on female labor supply indicates a conflict between market work and caring for children which is not resolved even though Norwegian institutions ease the combination of work and family, e.g. through rights to parental leave, rights to return to the same job after parental leave and good kindergarten coverage.

The last unconditional outcome is the probability of receiving out of work benefits. We find no significant effect of having the first, second or the third child. This is consistent with the previous findings that motherhood does not increase the probability of being marginally or not employed at all. There is a reduction in out of work benefits during pregnancy and when the child is between 0 and 1, which suggests that benefits like sickness absence and parental leave are in part replacing other benefits.

Sickness absence requires employment[20], and we therefore condition on the individual earning more than 1 BA when we measure the effect on sickness absence (the same definition of employment that we use in the second column). We find that the second and the third child actually lowers sickness absence rates. The effect is just as large when we condition on working full time[21]. There seems therefore to be no support for the view that the combination

[20] It is in principle possible to have sickness absence from unemployment. This is however not much used and is here ignored.

[21] Results available upon request

of work and family obligations may have negative effects on health. They are rather positive.

It is not the case, however, that effects of having children cannot contribute to understanding the gender-gap in sickness absence for employed Norwegians, which lies around 3 percentage points. The most important fertility outcome that affects sickness absence is pregnancy which increases sickness absence 36 percentage points. This may be interpreted as a causal effect of pregnancy, since pregnancy has no significant effect on earnings or labor supply. Using data from 2009, we do a back-of-the-envelope calculation of how much of the gender gap in sickness absence can be explained by sickness absence during pregnancy. The sickness absence rate for women was in that year 7.65 and 4.56 for men. With 4.83 percent pregnant women in the labor force and 35% higher sickness absence for pregnant women, the corresponding absence rate for non-pregnant women is 5.92. The gender difference in sickness absence if we take out the effect of pregnancy is therefore only 1.36 persentage points. This means that pregnancy can explain more than half the overall gender difference in sickness absence. Doing the same excercise for the age-group 18-45, pregnancies can explain more than 80% of the gender differences in sickness absence[22].

A direct effect of the miscarriage is captured in the coefficient θ. We do find a negative own effect on earnings, employment and sickness absence. A miscarriage therefore has negative labor market effects, most visibly reduced earnings. The negative effect on sickness absence is not straightforward to interpret, though, because employment is also affected directly. It might indicate lower sickness absence after a miscarriage, but it may also indicate that the women with the highest sickness absence are the ones who leave the labor market after the miscarriage. In section 7 we discuss the own effect further.

With earnings reduced by 18% and labor supply (total hours) reduced by 10%, a part of the effect on earnings can not be explained by a reduction in labor supply. This suggests a substantial wage penalty to motherhood of 8%. With the probability of underestimating the effect on labor supply, however, we cannot directly infer the size of the wage-penalty from these numbers. A wage penalty to motherhood has been put forward as a potential explanation

[22] Thses numbers are consistent with previous findings on Norwegian data (Rieck and Telle, 2012).

for parts of the male-to-female wage gap observed across all OECD countries, and we cannot outrule such an effect. The wage penalty indicates that women's careers lag behind when they have children.

6 Comparison to the samesex instrument

In this section, we estimate the effect of having a third child on labor market outcomes using the *samesex* instrument used by Angrist and Evans (1998) to compare our identification strategy to a (by now) standard strategy. The complier groups are different when using miscarriage and *samesex* as instruments, and results are not directly comparable, but they will provide an indication of whether the results are in the same range.

Using the *samesex* instrument, we are only able to estimate the effect of one endogenous variable - namely the probability of having a third child. It is also not possible to identify the exact effect of having a third child below five years old (which is the age of the children using miscarriage as instrument), because the random draw is when the *second* child is born. To get comparable results, we estimate the mean effect of having a third child on a sample who had their second child 3-7 years before in the years 2001-2004. In this sample, observed third children are between 0 and 7 years old. The share of women with a 7 and a 6 year old in the last year of observation is however small (0.0008% and 0.079% respectively). The results are therefore mainly identified by women who have a third child between 0 and 5 years old.

When using (miscarriage $\times$ g) as instruments, we were able to identify the effect of pregnancy, having an infant and having a toddler in addition to the marginal effect of having the third child. Using the samesex instrument, the effect of having a third child will capture a weighted mean of all these variables - which illustrates also an advantage of our modelling strategy in this paper: we get detailed estimates of the main variables defining motherhood in the first five years that are easy to interpret. To get a comparable mean from the results using (miscarriage $\times$ g), we need to add the results for being pregnant, having an infant and (having a toddler $\times$ 2) to the effect of (having a third child $\times$ 5) and divide by 5 to get the mean effect

Table 5: 2SLS estimates of the effect of having a third child on Employment, Earnings and Out of work benefits using *samesex* as instrument compared to the calculated mean effect of having a third child using coefficients from Table (4)

	samesex			*miscarriage* $\times g$		
	Empl	Earn	Benefits	Empl	Earn	Benefits
Third child	-0.063**	-11 061***	0.021	-0.058	-10 698	-0.017
	(0.031)	(4 136)	(0.036)			
N	97 721	97 721	97 721	401 955	401 955	401 955

Notes: included covariates in columns 1-3 are: year (4 dummy variables), nationality (5 dummy variables), years of schooling (9 categories), boy first, boy second (similar to Angrist and Evans (1998)). Each column is one regression.

of having the third child.

In Table (5), we report the results from a 2SLS estimation of the effect of having a third child using the *samesex* instrument, and the calculated mean effect of having a third child using (miscarriage $\times$ g) as instruments. We only use the significant coeffients from Table (4), hospital miscarriage sample. The number of observations is smaller when using the *samesex* instrument because we only use a sample who had two children in the period 2001-2004 to identify the effects.

The results are very similar when using the *samesex* instrument. Results using (miscarriage $\times$ g) as instruments can therefore cast light on what lies behind the effects using *samesex* as instrument - it maps a broad measure of "having a third child" into to the relative impact of being pregnant, having an infant, having a toddler and an older child. This is especially useful when estimating short-term effects of having children when the age of the child potentially has a larger impact than in the long term. It also tells us about the external validity of the impact of a third child to other parities - to the effect of the first and the second child.

Table 6: Robustness checks

	Earnings		Employment		Hours cond.	
	$g = t \times birth$	Exclude $t = 1$	$g = t \times birth$	Exclude $t = 1$	$g = t \times birth$	Exclude $t = 1$
First child	-9868*	-10960	-0.047	-0.060	-1.13	-1.82
	(5912)	(37359)	(0.064)	(0.380)	(1.88)	(11.76)
Second child	-10737*	-12612	-0.050	-0.043	-0.43	-2.02
	(5796)	(38117)	(0.065)	(0.387)	(1.93)	(13.00)
Third child	-11035**	-9186	-0.075	-0.049	-1.48	-1.88
	(5376)	(60354)	(0.058)	(0.613)	(1.79)	(20.42)
Pregnancy	4388	173	0.066	0.088	5.91	2.85
	(2137)	(359320)	(0.136)	(3.65)	(4.10)	(107.58)
Children 0-1	-378	-516	0.040***	0.039	0.40	0.86
	(1172)	(32179)	(0.015)	(0.327)	(0.59)	(11.42)
Children 1-3	-23142***	-2093	-0.025***	-0.022	-0.32	-0.097
	(578)	(17364)	(0.007)	(0.177)	(0.26)	(6.10)
θ	-2390	-2608**	-0.019	-0.017	0.39	-0.015
	(2363)	(1311)	(0.026)	(0.014)	(0,76)	(0,691)
N	401 955	321 564	401 955	321 564	369 889	293 082

Notes: included covariates are: z_i and g_{it}, year at planned birth (4 dummy variables), month at planned birth (12 dummy variables), nationality (5 dummy variables), years of schooling (9 categories) and two-digits sectoral industry codes (ISIC) at planned birth (61 dummy variables). Standard errors clustered on group g are reported in parenthesis. Each column is one regression.
Highest C.I. refers to the Belsley et al. (1980) test for near collinearity.

7 Robustness

We conduct three important robustness tests, displayed in Tables 6 and 7. We use the hospital sample to conduct the tests, but using the GP sample gives similar conclusions[23].

In our first robustness test we exclude age from the grouping variable g used to construct the instrumental variable. We thereby reduce the number of instruments considerably. If the coefficients change, this might be an indication of weak instruments and finite-sample bias as discussed in Section (2). Other reasons for not using the *age* dimension is that the estimated βs should not be sensitive to our choice of groups in the case of homogeneous treatment effects. In our application, the dimensions *birth number* and *time* are crucial for the identification of the marginal effect of each child and the time-varying variables pregnancy and the presence of small children. However, the dimension *age*, which also is used to construct the groups, is not as vital for the identification. There are also reasons to believe that the effect of children might

[23]Results available upon request.

vary over the life-cycle, making the homogeneous effects assumption too strict.

The results when age is excluded from g are displayed in columns 1,3 and 5 in Table 6. We see the standard errors grow considerably. The point estimates are close to the baseline estimates in Table 4, but with large standard error the results for employment and hours are not significant. This underlines the main reason for using the full instrument set in the main spesification. The results are almost unaltered indicating that our main specification does not suffer from finite-sample bias.

The second robustness check is to estimate the model without using the first year after planned birth. A potentially crucial assumption in our modeling strategy is that we assume the direct effect of having a miscarriage, if any, to be constant across groups and consequently constant over time. This is at odds with the findings that miscarriage is followed by a period of grief which wavers off after a period of 6-12 months ((Broen et al., 2004; Lok et al., 2010; Brier, 2008). In order to test this assumption we exclude the first year after birth/miscarriage from the estimation. The results are displayed in columns 2,4 and 6 in Table 6. Our results are also robust to this potential source of bias. The point estimates are very similar to the baseline specification, but the results are no longer significant because of the smaller sample size. The estimated own effect of a miscarriage (θ) is not affected much by excluding the first year, indicating that the constant effect assumption seems reasonable.

The third robustness test is to estimate the model separately for 1st-3rd births. The results are reported in Table 7. Also here, estimates are less significant due to fewer observations. Point estimates are however quite close to the baseline still. The conclusion from the baseline model that the marginal effect of each child is equally large is supported. The marginal effect of the first child estimated by first births is not significantly different from the marginal effect of the second child estimated by second births etc. Point estimates are even more equal for toal hours. The variables pregnant, having an infant and having a toddler are estimated jointly for 1st-3rd births in the baseline models. Estimating them separately, we get impresise estimates, but coefficients that are significant are not significantly different from eachother which is reassuring.

Table 7: 2SLS estimates using $g = t \times age$ as instruments. Estimating the model separately for 1st-3rd births

	Earnings			Total hours		
	1st birth	2nd birth	3rd birth	1st birth	2nd birth	3rd birth
First child	-11713***			-3.63**		
	(4115)			(1.60)		
Second child	-13145*	-15840***		-0.41	-3.04	
	(7533)	(5833)		(2.90)	(1.89)	
Third child	26942	-5160	-6370	-9.24	-5.63	-3.08
	(41990)	(10525)	(5657)	(16.53)	(4.21)	(2.52)
Pregnancy	7319	-12980	260	-2.84	2.06	5.79
	(7895)	(11477)	(10795)	(3.33)	(3.76)	(5.75)
Children 0-1	-2876**	-1401	-5310**	-0.94	0.93	0.52
	(1371)	(1969)	(2077)	(0.74)	(0.89)	(1.15)
Children 1-3	-23173*	-87	-3316***	-0.53	-0.22	-0.38
	(1321)	(1384)	(1267)	(0.58)	(0.55)	(0.63)
θ	-2552	-2614	-524	0.12	-0.21	-0.56
	(2424)	(1756)	(3017)	(0.95)	(0,75)	(1.37)
N	239 750	108 935	53 270	239 750	108935	53 270

Each column is a 2SLS regression. Control variables not reported in the table are year at planned birth (4 dummy variables), month at planned birth (12 dummy variables), nationality (5 dummy variables), years of schooling (9 categories) and two-digits sectoral industry codes (ISIC) at planned birth (61 dummy variables). Standard errors clustered on group g are reported in parenthesis. Each column is one regression.
Highest C.I. refers to the Belsley et al. (1980) test for near collinearity.

8 Conclusion

Childbearing and -caring is an important source of inequality in labor markets. Many countries have established institutions designed to ease the combination of motherhood and employment, such as (nearly) universal and highly subsidized child care and fully paid parental leave for almost a year. Still we observe substantial differences in labor market outcomes both between men and women as well as between women with and without children.

In this paper we have estimated the effect of motherhood on female labor market outcomes using a biological fertility shock; miscarriage. *Motherhood* is modeled as a combination of longer term and temporary effects related to having small children, in total six fertility outcomes. With several endogenous variables, we also needed multiple instruments. There are few available instruments for fertility outcomes, and we increase the number of instruments by exploiting heterogeneity in the treatment response across different dimensions.

We find that each of the first children has a large impact on female earnings, which on average are reduced by 18%. The largest decrease in earnings are driven by women with average earnings reducing their labor supply and thereby earn less. There are small effects on earnings for those who are marginally employed or women in the higher end of the earnings distribution. Labor force participation is not affected much by having children, but the intensity of market work is reduced. Women work on average 2 hours less per week per child, and the effect does not decrease (much) when the child grows older. Motherhood therefore plays a significant role in explaining female part-time work. We find no evidence of an adverse health effect of having children as neither sickness absence nor disability seem to increase due to motherhood.

Despite one of the most extensive set of labor market policies to support mothers' employment in the world, entering motherhood substantially reduces Norwegian women's earnings as well as their hours worked. Our data does not tell us whether this is due to preferences for spending time with their children or an undesired consequence of an increased work burden at home. However, the complete absence of evidence for adverse health effects, both measured as sickness absence and uptake of other benefits (including disability), points towards observed labor market responses to motherhood being mostly due to preferences for reduced labor supply.

References

Agüero, J. M. and Marks, M. S. (2008). Motherhood and female labor force participation: Evidence from infertility shocks. *The American Economic Review*, 98(2):pp. 500–504.

Agüero, J. M. and Marks, M. S. (2011). Motherhood and female labor supply in the developing world: Evidence from infertility shocks. *Journal of Human Resources*, 46(4):800–826.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3):pp. 450–477.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):pp. 979–1014.

Ashcraft, A., Fernández-Val, I., and Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal*, 123(571):875–905.

Ashcraft, A. and Lang, K. (2006). The consequences of teenage childbearing. Working Paper 12485, National Bureau of Economic Research.

Belsley, D. (1991). A guide to using the collinearity diagnostics. *Computational Economics*, 4(1):33–50.

Belsley, D., Kuh, E., and Welsch, R. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. *Wiley, New York*.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Bratberg, E., Dahl, S.-g., and Risa, A. E. (2002). ”the double burden”: Do combinations of career and family obligations increase sickness absence among women? *European Sociological Review*, 18(2):233–249.

Brier, N. (2008). Grief following miscarriage: A comprehensive review of the literature. *Journal of Women's Health*, 17(3):451–464.

Broen, A. N., Moum, T., Bødtker, A. S., and Ekeberg, i. (2004). Psychological impact on women of miscarriage versus induced abortion: A 2-year follow-up study. *Psychosomatic Medicine*, 66(2):265–271.

Buckles, K. S. and Munnich, E. L. (2012). Birth spacing and sibling outcomes. *Journal of Human Resources*, 47(3):613–642.

Budig, M. J. and England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, 66(2):204–225.

Cristia, J. P. (2008). The effect of a first child on female labor supply: Evidence from women seeking fertility services. *Journal of Human Resources*, 43(3):487–510.

Garcia-Enguidanos, A., Calle, M., Valero, J., Luna, S., and Dominguez-Rojas, V. (2002). Risk factors in miscarriage: a review. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 102(2):111–119.

Goldin, C. (1990). *Understanding the Gender Gap: An Economic History of American Women*. Number gold90-1 in NBER Books. New York: Oxford University Press.

Harkness, S. and Waldfogel, J. (2003). *The Family Gap in Pay: Evidence from Seven Industrialized Countries*, volume 22 of *Worker Well-Being and Public Policy (Research in Labour Economics)*, pages 369–413. Emerald Group Publishing Limited.

Hotz, V. J., McElroy, S. W., and Sanders, S. G. (2005). Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *The Journal of Human Resources*, 40(3):pp. 683–715.

Hotz, V. J., Mullin, C. H., and Sanders, S. G. (1997). Bounding causal effects using data from a contaminated natural experiment: Analysing the effects of teenage childbearing. *Review of Economic Studies*, 64(4):575–603.

Åkerlind, I., Alexanderson, K., Hensing, G., Leijon, M., and Bjurulf, P. (1996). Sex differences in sickness absence in relation to parental status. *Scandinavian Journal of Public Health*, 24(1):27–35.

Kline, J., Susser, M., and Stein, Z. (1989). *Conception to birth : epidemiology of prenatal development / Jennie Kline, Zena Stein, Mervyn Susser*. Oxford University Press, New York.

Kolesár, M., Chetty, R., Friedman, J. N., Glaeser, E. L., and Imbens, G. W. (2011). Identification and inference with many invalid instruments. Working Paper 17519, National Bureau of Economic Research.

Korenman, S. and Neumark, D. (1992). Marriage, motherhood, and wages. *Journal of Human Resources*, 27(2):233–255.

Lok, I. H., Yip, A. S.-K., Lee, D. T.-S., Sahota, D., and Chung, T. K.-H. (2010). A 1-year longitudinal study of psychological morbidity after miscarriage. *Fertility and Sterility*, 93(6):1966–1975.

Mastekaasa, A. (2000). Parenthood, gender and sickness absence. *Social Science and Medicine*, 50(12):1827–1842.

Miller, A. (2011). The effects of motherhood timing on career path. *Journal of Population Economics*, 24(3):1071–1100.

Mincer, J. and Polachek, S. (1974). Family investment in human capital: Earnings of women. *Journal of Political Economy*, 82(2):S76–S108.

Rieck, K. M. E. and Telle, K. (2012). Sick leave before, during and after pregnancy. Working Papers in Economics 06/12, University of Bergen, Department of Economics.

Rosenzweig, M. R. and Wolpin, K. I. (1980). Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 88(2):pp. 328–348.

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.

Wang, X., Chen, C., Wang, L., Chen, D., Guang, W., and French, J. (2003). Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertility and Sterility*, 79(3):577–584.

Wilcox, A. J., Weinberg, C. R., O'Connor, J. F., Baird, D. D., Schlatterer, J. P., Canfield, R. E., Armstrong, E. G., and Nisula, B. C. (1988). Incidence of early loss of pregnancy. *The New England Journal of Medicine*, 319(4):189–194.

A Appendix

Table A.1: Descriptive statistics of the outcome variables

	Birth	Miscarriage, GP	Miscarriage, hospital
First child			
Earnings t+1	42,247	53,602	59,464
Earnings t+5	54,262	50,556	56,649
Hours t+1	29.1	31.3	32.8
Hours t+5	29.7	29.1	30.4
Sickn.abs t+1	0.8	14.0	13.2
Sickn.abs t+5	5.8	7.4	6.5
Disability/welfare t+1	1.6	3.3	1.9
Disability/welfare t+5	4.8	6.4	4.2
Second child			
Earnings t+1	46,717	52,265	59,222
Earnings t+5	55,145	53,851	59,196
Hours t+1	27.3	29.1	30.8
Hours t+5	29.3	29.3	30.9
Sickn.abs t+1	0.8	15.1	16.9
Sickn.abs t+5	5.9	7.8	4.6
Disability/welfare t+1	1.7	3.7	1.8
Disability/welfare t+5	6.9	7.3	7.4
Third child			
Earnings t+1	44,264	51,291	57,484
Earnings t+5	54,986	52,548	59,470
Hours t+1	25.2	28.3	29.5
Hours t+5	28.2	28.1	29.5
Sickn.abs t+1	0.9	14.6	11.8
Sickn.abs t+5	5.9	6.8	6.4
Disability/welfare t+1	1.5	3.7	3.2
Disability/welfare t+5	7.2	9.8	8.4

Earnings are both labor and capital incomes. Work hours are set to 0 if the individual is not employed and is measured as the average working week in a year. Sickness benefits are welfare benefits for sickness periods longer than 16 days. Other benefits are various social welfare benefits that all citizens are eligible for, regardless of employment.

Table A.2: 2SLS estimates of the main model (Equation 2 and 3): The impact of six fertility outcomes on employment, earnings and out of work benefits. GP sample

	Unconditional outcomes				Cond. on employment	
	Earnings	Employment	Hours	Benefits	Hours	Sickn. abs.
First child	-8 507***	-0.093***	-5.22***	0.005	-2.88***	0.00
	(2 485)	(0.027)	(1.10)	(0.021)	(0.89)	(0.02)
Second child	-6 449***	-0.083***	-3.73***	-0.005	-1.59*	-0.01
	(2 528)	(0.028)	(1.16)	(0.024)	(0.95)	(0.02)
Third child	-7 469***	-0.085***	-4.65***	-0.007	-2.34	-0.02
	(2 512)	(0.029)	(1.13)	(0.025)	(0.93)	(0.02)
Pregnancy	-2 558	-0.063	-3.15	-0.081**	-2.63*	0.31***
	(4 421)	(0.047)	(2.08)	(0.035)	(1.81)	(0.03)
Children 0-1	-3 449***	0.027*	-0.32	-0.044***	-1.62***	-0.01
	(1 005)	(0.014)	(0.58)	(0.012)	(0.54)	(0.01)
Children 1-3	-2 159***	-0.014*	-0.69**	0.008	-0.59*	-0.05***
	(565)	(0.008)	(0.31)	(0.006)	(0.28)	(0.01)
θ	-2 564**	-0.024**	-1.26**	0.004	-0.54	0.01
	(1 064)	(0.012)	(0.49)	(0.010)	(0.40)	(0.01)
Highest C.I.	11.81	11.81	11.81	11.81	11.7	11.7
J-value, overidentification	260.6	259.7		378.0	408.0	341.7
p-value, overidentification	0.998	0.998		0.032	0.002	0.303
N	402 365	402 365	402 365	402 365	370 168	370 168

Notes: included covariates are: z_i and g_{it}, year at planned birth (4 dummy variables), month at planned birth (12 dummy variables), nationality (5 dummy variables), years of schooling (9 categories) and two-digits sectoral industry codes (ISIC) at planned birth (61 dummy variables). Standard errors clustered on group g are reported in parenthesis. Each column is one regression.
Highest C.I. refers to the Belsley et al. (1980) test for near collinearity.