HAPPY DOCTOR MAKES HAPPY BABY? INCENTIVIZING PHYSICIANS IMPROVES QUALITY OF PRENATAL CARE

Vibeke Myrup Jensen*

Abstract—Physician-induced demand, whereby physicians alter patient treatment for personal gain, lies at the heart of concerns about publicly provided health care. However, little is known about how payment systems affect the ultimate outcome of patient health. Exploiting a unique policy induced variation in Denmark, I investigate the impact of physician payment contracts on infant health. In a difference-in-differences framework, I find that firstborn infants exposed in the womb to the care of general practitioners with capitation contracts have poorer infant health outcomes than infants exposed to fee-for-service contracts. The firstborn children of younger women primarily drive the effects.

1. Introduction

Physician-induced demand, whereby physicians alter their patient treatment in the pursuit of personal gain, is not only a central issue in public health care programs but also lies at the heart of concerns about publicly provided health care such as Medicare. While several studies investigating the effects of physician payment systems find that economic incentives increase the number of services provided (Gruber & Owings, 1996; Mullen, Frank, & Rosenthal, 2010), no studies investigate how economic incentives affect the ultimate outcome: the health patients. This paper fills a gap in this literature by estimating the effect of policy-induced changes in the physician remuneration system on infant health.

Physician behavior is a central issue in any health system because of the asymmetric information between the physician and the patient. The literature covers the relationship between financial incentives and medical practice as a classical income-leisure model (McGuire, 2000; Woodward & Warren-Boulton, 1984). When the physician maximizes his or her utility function, this physician is, at the margin, willing to surrender some accuracy about the treatment options in the interest of personal gain.

To identify the effects of remuneration changes on newborns, I exploit an influential natural experiment in Denmark. In 1987, general practitioners (GPs) in Copenhagen moved from pure capitation to a mix of one-third capitation and two-thirds fee-for-service (FFS) contracts, bringing them into line with contracts in the rest of Denmark. Thus, as GPs on capitation contracts could maximize their personal gain only by increasing leisure time and GPs on FFS contracts could maximize their personal gain either by increasing the amount of patient treatment or by shifting to more expensive types of treatment (Scott, 2000), the economic incentives changed for GPs in Copenhagen.

Denmark is ideally suited for such a natural experiment. First, selection into treatment is not an issue, because of universal mandatory health insurance coverage and a GP listing system based on patient residential address. Second, no recall bias exists in the data, as I use high-quality administrative data where birth records link directly to parental background information.

Krasnik et al. (1990) estimate the effect of the Copenhagen reform on practice behavior, finding that the reform increases the self-provision of service by reduced referral rates (to specialists and hospitals) and an increased number of diagnostic and curative services. For example, they find that diagnostic services rose by 51.9 services per 1,000 patients. While they provide evidence that the reform had an effect on GP behavior, a key question is how the reform affected the patients.

Specifically, I estimate a difference-in-differences model (DD). In Denmark, pregnant women receive medical treatment mainly from their GP. For a clean distinction between exposures to different types of payment systems, the treatment of interest is infants exposed in the womb to the care of GPs with capitation contracts compared to infants exposed in the womb to the care of GPs with partial FFS contracts. Moreover, to eliminate investment behavior related to previous pregnancy experiences, I restrict my sample to firstborns. Thus, I define firstborn infants in the City of Copenhagen as the treatment group and firstborn infants in the geographically adjacent municipalities surrounding the city, Copenhagen County, as the control group.

In line with previous medical studies, I find that the probability of poor pregnancy outcomes is U-shaped, with the worst outcomes for adolescent and women of advanced maternal age. Thus, I investigate the effects of capitation contracts on younger (under age 27), prime age (ages 27–30), and older women (above age 30). To the extent that the only important difference in county policies is the remuneration system, I find that particularly infants born to younger women treated by GPs on capitation contracts during their pregnancy have worse health outcomes than those born to younger women treated by GPs on partly FFS contracts. For the younger women, I find a 1.8% lower birth weight, a 58% higher probability of preterm birth (before...
the 37th week), and a 1.2% slower fetal growth rate under capitation contracts than under partly FFS contracts. However, I find small and imprecisely estimated effects on low birth weight (below 2,500 gram) and very preterm births (before the 34th week). While for prime-age and older women I find no significant effects, further investigations of these age effects indicate a U-shaped effect of capitation contracts on birth weight, with the largest effects for younger and older women.

To validate the results, I add two steps. First, I investigate how balanced the treatment and control groups are by estimating an unconditional DD model for each control variable. Second, I apply a robustness check: estimating the DD model using data before the reform. In general, the results pass this test.

The medical literature suggests that the underlying causes of poor pregnancy outcomes are different for younger and older women. While, for example, advanced maternal age directly increases the probability of preterm births, a higher rate of risk behavior during pregnancy predicts preterm births for younger women (Ohlsson & Shah, 2008). In addition, older and younger women have different levels of education and thus different levels of bargaining power. Therefore, selection into different treatment patterns potentially explains why I find significant effects for only the younger women. Three pieces of empirical evidence support this theory. First, I find that the number of all types of specialist visits during pregnancy increases by age, whereas the number of GP visits decreases by age. Second, using the level of schooling as a proxy for risk behavior, I find that the change in remuneration system generally has the smallest effect on women with a college degree. Third, younger women with only a basic level of schooling, the highest level of risk behavior, drive the effects found for the younger women.

A limitation of the study is that the increase in GP services directly related to prenatal care is unknown. This limitation also complicates cost-efficiency calculations of the reform. As an alternative, to contextualize my findings, I first compare them to the effects of maternal schooling on birth weight. For younger women, I find that the birth weight improvement of incentivizing GPs falls between half the birth weight increase from one additional year of maternal schooling for the United Kingdom (Chevalier & O’Sullivan, 2007), two-thirds the increase for the United States (Currie & Moretti, 2003), and double the increase for Denmark (Bingley, Christensen, & Jensen, 2009). Second, to frame a possible spillover, I compare my birth weight results to the existing studies of long-run effects of birth weight. I find that the effects of changing the remuneration system would be a 1.5% increase in high school completion (Black, Devereux, & Salvanes, 2007) and 0.08 additional months of schooling (Royer, 2009).

The paper proceeds as follows. Section II outlines the background information and the conceptual framework. Section III describes the data. Section IV presents the results, and section V concludes and contextualizes the findings.

II. Background and Conceptual Framework

A. The 1987 Change in the GP Remuneration System

On October 1, 1987, following national discussions about workload differences—and thereby potential income differences—between GPs in different counties, the Danish National Board of Health unified GP remuneration systems. GPs in the municipalities of Copenhagen City then shifted from pure capitation contracts to the standard mix of one-third capitation and two-thirds FFS contracts found elsewhere in Denmark.

The 1987 change in remuneration system directly increased the number of services that the GPs in Copenhagen City performed. Krasnik et al. (1990) investigate the effects of the change in GP remuneration system on the level of GP services, six months before, at six months, and twelve months after October 1987. They randomly selected 100 of 265 GPs in Copenhagen City and collected information about the GPs and the services they provided to their 125,000 patients. As controls, Krasnik et al. use county-level billing fees and referrals from all 365 GPs in Copenhagen County for the same three periods.

Table 1 lists the main findings from Krasnik et al. (1990). The table shows the estimated changes in the number of services provided per 1,000 patients for both Copenhagen City and Copenhagen County six months before and six months after the remuneration system changed. To this table, I add an unconditional DD estimator for each type of contact or service, allowing me to compare Copenhagen City and Copenhagen County before and after the reform.

This DD estimator shows that the change in the remuneration system affected both the rate of GP contacts and the rate of GP activities. For example, telephone consultations rose by 11.4 contacts per 1,000 enlisted patients, and

1 A change from the base of 5.3% preterm births.
2 For the older women, the effects of capitation contracts on the different levels of education are less systematic than for the younger women.

3 In the 1980s, GP per patient workload increased due to earlier discharges from hospitals and increases in both waiting times for hospitalization and the proportion of elderly patients (Krasnik et al., 1990).
4 Unfortunately, there are no exact data available on GP income in the 1980s; until 1992, income information for GPs was pooled with doctors working at hospitals, dentists, and veterinarians.
5 Seventy-one of the 100 GPs responded with full information for all three periods.
6 As no public information on GP services is available until 1992, I rely on the estimates from Krasnik et al. (1990).
curative care increased by 79.8 contacts per 1,000 enlisted patients, while the number of referrals to specialists and hospitals fell by 21.1 and 33.7 referrals per 1,000 enlisted patients, respectively.7

Although not directly targeting service in prenatal care, this study provides some evidence that GPs in general acted on the reform. FFS, which makes up two-thirds of the GP remuneration system, includes special fees for 4 types of consultation, 39 types of medical services, and 37 types of laboratory tests, including standard services during pregnancy such as GP visits and urine and blood tests for pregnancy-related diabetes (Praktiserende Laegers Organisation, 1984). As the standard type of services for pregnant women is included in the list of services that are reimbursed in the FFS system, I assume that the general change in GP behavior also affected their treatment of pregnant women.

B. Defining the Treatment and Control Areas

Given the natural experiment setting in which some areas were exposed to remuneration changes and others were not, I estimate the effects of the GP payment system on infant health. Following Krasnik et al. (1990), I define Copenhagen City as the treatment area and Copenhagen County as the control area. While Copenhagen City and County are separate health markets, each contains minor administrative units: municipalities. Municipalities are local authorities in terms of several types of social services (such as primary schools), and each municipality has different tax levels. The treatment area consists of four municipalities, and Copenhagen County comprises sixteen municipalities surrounding Copenhagen City. I estimate the effects using infants conceived and born during a three-year period before (October 1984–September 1987) and a three-year period after (October 1987–September 1990) the remuneration system changed in Copenhagen City.8

C. Prenatal Care

The treatment and control groups constitute women exposed to prenatal care by GPs under different payment systems. This section provides information about standard prenatal care in Denmark and considers potential differences in prenatal care between health care markets.

In Denmark pregnant women mainly consult their GP or midwives. In prenatal care (as in other types of health services), the Danish GP acts as a gatekeeper to all types of nonemergency health care (Vallgaarda, Krasnik, & Vrangbaek, 2001). Regular prenatal care by GPs and midwives has been the standard service in Denmark since 1945 (Vedsted et al., 2005). These consultations are regulated by national law, guaranteeing three free GP consultations, five free midwife consultations, and free midwifery. No major legal changes that could have offset differences in prenatal care affected either the treatment or the control group (Sundhedsstyrelsen, 1985, 1998).

Upon suspecting that she is pregnant, a woman visits her GP, who determines the due date if she is pregnant. The second and third GP visits take place during the second and third trimesters. These three visits are the minimum requirements for the GP to provide. If the fetus is at any risk, the GP can ask the woman to visit more regularly or refer her to an obstetrician (OB).9 Referrals also occur if the woman has a higher probability of a poor pregnancy outcome. For example, national recommendations suggest ultrasound screenings for Down syndrome for women over 35 years old or with a family history of this syndrome (Sundhedsstyrelsen, 2004b).

The prenatal care consultations that the GP provides consist of tests and counseling. The national recommendations during the 1980s included urine test, blood tests, and counseling in terms of “appropriate lifestyle” during pregnancy. Testing is important for preventing and curing infections that could otherwise lead to premature births. Consultations about diet, consumption of alcohol, and smoking are helpful in ensuring sufficient nutrition for the fetus (Sundhedsstyrelsen, 1985). As all of these tests and face-to-face and telephone

---

7 The reduction in specialist treatments does not necessarily imply a reduction in the quality of care. For example, the standard physical examination before prescribing contraception is within the standard scope of GP care. Alternatively, the GP can make a referral to a gynecologist.

8 In 1987, the average population per municipality was 152,150 individuals in the treatment area and 34,658 individuals in the control area.

9 These OB visits are at the hospital.
consultations changed from being part of the capitation contracts to FFS in 1987 (Praktiserende Laegers Organisation, 1984), the level of services was likely to change with the reform.

At the first visit, the GP refers the woman to a group of midwives, who invite her to two consultations during the second trimester and three during the third. These consultations consist primarily of general monitoring of the pregnancy and of information about the different stages of pregnancy and lifestyle during pregnancy. If the midwives suspect a problem, they encourage the woman to visit her GP (Sundhedsstyrelsen, 1985).

In addition to these GP and midwife consultations, some counties provide ultrasound screenings. Thus, a potential bias exists if the implementation of obstetric ultrasound screenings happens more quickly in one area than in another. Because no data exist on the coverage of ultrasound screening before the GP reform, investigating the impact of this potential bias is impossible. Nonetheless, in 1990, the rate of women having at least one obstetric ultrasound screening and the average number of obstetric ultrasound screenings per pregnancy was similar for the treatment area (94.6% and 2.2 screenings) and the control area (93.5% and 1.7 screenings) (Joergensen, 1993, 2003).

D. Potential Biases from Shifts in GPs or Patient Group Composition

The change in remuneration system in Copenhagen City constitutes a well-defined change in GP payment systems. However, as Copenhagen City likely became more attractive to GPs after the implementation of FFS contracts, two biases potentially contaminate the distinction between the treatment and control areas. First, the implementation of FFS contracts could alter the influx of GPs to Copenhagen City. Second, the shift in GPs’ incentives for patient treatment could lead to a change in patient composition.

In Denmark, while the counties are the local health authorities, the central government tries to maintain national homogeneity in health care through block grants and national regulations. Some regulations limit rapid movements between health care markets, such as the regulation of the practice location through the GP-to-patient ratio. This regulation specifies that GPs can change location only if the GP-to-patient ratio exceeds a certain limit in that neighborhood. However, while GP quality between the treatment and control group changes slowly, potential changes in the local tax level create an immediate change in GP incentives to work. Nonetheless, as the positive trend in the local tax level for GPs is identical for the two areas throughout the period, I find no such dilution of the effects (see online appendix figure A1).

A different threat to the identification strategy lies in potential changes in patient group composition as a consequence of changes in GP incentives. Fortunately, GPs in Denmark have little chance of cream-skimming patients, for two reasons First, health care insurance in Denmark is publicly provided, so that different health insurance plans do not exist and therefore cannot affect patient or treatment choices. Second, residential address defines GP enrollment. With a minimum of two GPs per neighborhood, the patient chooses the GP from publicly available information about practice size, GP gender, and age. Children are automatically assigned to their mother’s GP, and the possibility of reallocation within a municipality without moving occurs only annually.

E. Identification Strategy

I evaluate the impact of changes in the GP remuneration system on infant health by estimating a DD model. The key estimation for infant $i$ in municipality $m$ at time $t$ is

$$Health_{itm} = \beta_{0m} + \beta_1(Treat\_area_m \times Cap_t) + \beta_2Cap_t + \beta_3Year_t + X_{itm}b_4 + e_{itm},$$

where $Health_{itm}$ represents different measures of infant health. $\beta_{0m}$ defines municipality-fixed effects, which encompass any time constant effects common to municipalities. $Treat\_area_m$ is a binary variable equal to 1 for Copenhagen City and 0 in Copenhagen County. $Cap_t$ is a binary variable, where one defines the period during which GPs in Copenhagen City were on capitation contracts, and 0 defines the period after mixed contracts of capitation and FFS were implemented in Copenhagen City. The interaction between $Treat\_area_m$ and $Cap_t$ defines the variable of interest. $Year_t$ is a set of birth year and birth month dummies to account for a time trend. $X_{itm}$ denotes a vector of individual and parental characteristics, and $e_{itm}$ is the individual-specific error term. I use OLS to estimate equation (1) and compute standard errors clustered by municipality. I define the level of significance to 5% and the level of marginal significance to 10%.

III. Data

A. Data

The data set uses a sample of infants conceived and born between October 1981 and September 1990 in Copenhagen.
Table 2.—Summary Statistics for the Outcome Variables

<table>
<thead>
<tr>
<th>Full sample</th>
<th>Treatment Group</th>
<th>Control Group</th>
<th>DD Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth weight</td>
<td>Pre</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>(grams)</td>
<td>3,314.6</td>
<td>3,365.6</td>
<td>3,330.3</td>
</tr>
<tr>
<td>Low birth weight</td>
<td>0.0580</td>
<td>0.0510</td>
<td>0.0570</td>
</tr>
<tr>
<td>(&lt;2,500 grams)</td>
<td>(0.234)</td>
<td>(0.219)</td>
<td>(0.233)</td>
</tr>
<tr>
<td>Preterm birth</td>
<td>0.0570</td>
<td>0.0490</td>
<td>0.0470</td>
</tr>
<tr>
<td>(&lt;37th week)</td>
<td>(0.231)</td>
<td>(0.216)</td>
<td>(0.211)</td>
</tr>
<tr>
<td>Very preterm birth</td>
<td>0.0070</td>
<td>0.0070</td>
<td>0.0050</td>
</tr>
<tr>
<td>(&lt;34th week)</td>
<td>(0.084)</td>
<td>(0.085)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>Fetal growth</td>
<td>83.53</td>
<td>84.77</td>
<td>84.03</td>
</tr>
<tr>
<td>(grams/week)</td>
<td>(12.212)</td>
<td>(12.286)</td>
<td>(12.056)</td>
</tr>
<tr>
<td>Observations</td>
<td>7,054</td>
<td>9,043</td>
<td>5,434</td>
</tr>
</tbody>
</table>

For each outcome, the table reports the means, standard deviations (in parentheses), and number of observations separately for the periods before and after the reform. The prereform period defines infants conceived and born October 1984 to September 1987. The postreform period defines infants conceived and born October 1987 to September 1990. Because both types of remuneration system can affect infants born October 1987 to June 1988, I drop those. In addition to the means, the last column reports the unconditional DD estimator and robust standard errors (in brackets) between the treatment and control groups. I use a municipality fixed-effects model adjusted for birth year and birth month dummies to estimate the unconditional DD estimator. Significant at ***99%, **95%, *90%.

14 From 1982 to 1990, the average birth cohort in Denmark consisted of 58,957 live births per year.
15 I also try to use log gestational age; however, the effect of capitation contracts on this outcome is close to 0 and not significant throughout all estimations.
16 While the medical literature agrees on the U-shaped curve, there is no consensus on the division of maternal age. Thus, I use the age groups that fit the distribution of the data for low birth weight and preterm births.
17 However, a postreform increase in the percentage of preterm births in the control group partly explains this increase.
18 As in the full sample, the effects on preterm births are partly driven by a postreform increase in unhealthy infants in the control group.
19 The graph runs quarterly from the fourth birth quarter 1984 through third birth quarter 1990, and the overall and group-specific time trends are estimated as predicted means from the base of a municipality-fixed-effects regression including the full set of controls.
points. First, for most outcomes, the means are lower for the treatment group than for the control group in the prereform period, whereas there are very small differences or reversed differences in the postreform period. Second, in general, there are larger differences between the treatment and the control group for the younger mothers. However, for both groups and in the prereform period, larger differences in the mean outcomes exist between the treatment and the control groups in the earliest birth years than in those closer to the reform.

In summary, the graph indicates negative effects of capitation contracts. However, the means are decreasing over
time, and to some extent, the treatment and the control group do not follow the same pattern over time, indicating that time trends potentially affect the results. I attempt to take these time trends into account, first by investigating how balanced the treatment and the control group are in the set of covariates and by applying two robustness checks in section IVC.

B. Treatment and Control Group Balancing

To substantiate the causal interpretation of the estimates, I determine that the control and treatment groups balance in a set of background information.20 For the set of controls, table 3 provides summary statistics and the unconditional DD estimator for the full sample, and online appendix table A2 shows similar statistics separately for the three age groups. For the full sample, out of the eighteen control variables, three variables change concurrently with the reform, potentially biasing the results: women’s age when giving birth, percentage of women outside the labor force, and GP gender. I take the maternal age effects into account by estimating all results by different age groups and investigate the GP effect by running all regressions with and without controls for GP age, gender, and tenure. For the effects of maternal labor force participation, I run a separate regression without this variable. This exclusion does not change the results.

Online appendix table A2 shows, for the sample of older women, no significant changes in the covariates in the treatment group compared to that of the control group. For the sample of younger women, only GP gender changes significantly with the reform.21 For the prime-age women, the treatment and control groups do not balance for father’s log income, education, and the variables indicating unknown information on the father’s income and labor force participation. However, the results change only marginally when I exclude these three variables from the model.

20 I measure all parental background information in the year prior to giving birth.

21 A higher rate of male GPs in the postreform period for the treatment group drives this significant change.
TABLE 4.—EFFECT OF GP CAPITATION CONTRACTS ON INFANT HEALTH

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Younger Mothers</th>
<th>Prime-Age Mothers</th>
<th>Older Mothers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Log birth weight</td>
<td>−0.013*</td>
<td>−0.010</td>
<td>−0.018***</td>
<td>−0.018**</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Low birth weight</td>
<td>0.004</td>
<td>0.000</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>(&lt;2500 gram)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Preterm birth</td>
<td>0.016**</td>
<td>0.019**</td>
<td>0.022***</td>
<td>0.031***</td>
</tr>
<tr>
<td>(&lt;37th week)</td>
<td>(0.006)</td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Very preterm birth</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003*</td>
<td>0.003*</td>
</tr>
<tr>
<td>(&lt;34th week)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Fetal growth (grams/week)</td>
<td>−0.961**</td>
<td>−0.769**</td>
<td>−1.280***</td>
<td>−1.183**</td>
</tr>
<tr>
<td>(grams/week)</td>
<td>(0.403)</td>
<td>(0.409)</td>
<td>(0.382)</td>
<td>(0.439)</td>
</tr>
<tr>
<td>Control for GP</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>27,629</td>
<td>27,629</td>
<td>14,379</td>
<td>14,379</td>
</tr>
</tbody>
</table>

Each estimate represents a separate DD regression with robust standard errors in parentheses. October 1984–September 1987 defines the prereform period and July 1988–September 1990 the postreform period. Because infants born October 1987 to May 1988 are affected by both periods, I drop those. Table 3 lists the controls that are included in the regression. Significant at ***99%, **95%, *90%.

A. Estimating the Effects of Remuneration System on Infant Health

In this section, I estimate the effects of capitation contracts on infant health compared to the mixed system of capitation and FFS contracts for first-time mothers. Table 4 shows the effects of remuneration system changes on infant health for the full sample and subsamples of younger, prime-age, and older women, respectively. Columns 1 and 2 report the full sample for two sets of controls: infant characteristics and parental background alone and infant characteristics, parental background, and GP characteristics. Columns 3 and 4 report outcomes for younger women, columns 5 and 6 outcomes for prime-age women, and columns 7 and 8 outcomes for older women.

Estimating the results on the full sample and with the restricted set of controls, I find that birth weight is 1.3% lower under the capitation system. This result, however, is only marginally significant, and it is imprecisely estimated when I add the full set of controls. When I use the model that includes all controls, infants have a 1.9 percentage point (36.5% from the base of 5.2% births) higher probability of preterm birth. For low birth weight and very preterm birth, the effects are close to 0 and not significant. As low birth weight is a likely cause of preterm birth, I also estimate the effect on fetal growth. The infants born under capitation contracts have 0.8 grams per week, or 1% slower fetal growth rate. This effect, however, is only marginally significant.

For younger women, I find that infants born under the capitation contracts have on average 1.8% lower birth weight. These women also have a 2.2 percentage point (42% from the base of 5.3% births) higher probability of preterm birth, increasing to 3.1 percentage points (58% from the base of 5.3% births) when I control for GP information. Accounting for the interaction between birth weight and gestation, the fetal growth rate is −1.2 grams per week, or a 1.4% slower growth rate when compared to infants born under the mixed FFS and capitation contracts. As in the full sample, I find no significant effects on the probability of low birth weight and very preterm birth.

For the prime-age and older women, I find that most effects are imprecisely estimated, smaller than these effects found for the younger women.

For younger women I find effects of capitation contracts on average birth weight but not on low birth weight (below 2,500 grams). Therefore, I estimate the effects on different birth weight dummies (in 250 gram intervals) for this group, prime-age, and older women (online appendix table A5). For younger women, I mainly find that the marginal group of infants who benefit from the reform have a birth weight of between 3,000 to 3,250 grams. In contrast, for older women, I find significant effects of capitation contracts on the probability of births below 4,000 grams. As I mainly find effects of capitation contracts for women who were not at risk of very poor health outcomes, this finding suggests that women at risk were likely to be selected into specialist treatment in both periods.

Throughout the paper, I divide maternal age into three groups. For further investigations of this age effect, online appendix figure A3 graphs the effects of capitation contracts on log birth weight at different maternal ages. Although most estimates are not significant, the graph indicates the expected U-shape, with the largest negative effects

22 The restriction to firstborns implies a limitation because generalizing the results to all births is questionable. Online appendix table A3 shows the main results for all women and by parity (first-time mothers and mothers with at least one child). The table suggests that the effects for first-time mothers drive the effects found for all women. One exception is preterm birth, where higher-parity women drive the effect for all women.

23 While the table excludes the controls, online appendix table A4 presents the effect on log birth weight with all controls. In general, the coefficients have the expected signs, and several coefficients show minor significant effects, given the large sample size.

24 Some effects also have reversed signs.

25 Each estimate defines a separate regression using a six-year age interval. I loop through ages 16 to 35, that is, the estimate for age 16 equals the age group 16–21.
of capitation contracts for the younger and the older women.

B. Potential Explanations of the Heterogeneous Effects of Capitation Contracts

The medical literature suggests increased risks of poor infant health outcomes for both adolescent and women of advanced maternal age. However, the underlying causes of these risks are different for the two groups. Pregnancy-related risks can be a direct consequence of advanced maternal age, while only a few biological explanations exist for younger women (Jacobsson, Ladfors, Milsom, 2004; Ohlsson & Shah, 2008). Instead, for younger women, social and environmental factors have a higher impact on the fetus (Restrepo-Méndez et al., 2011). Several studies support this finding. Younger women have a higher probability of preterm births and low-birth-weight babies because of delayed prenatal care and higher levels of risk behavior such as alcohol consumption and smoking during pregnancy (Bateman & Simpson, 2005; Friede et al., 1987; Ohlsson & Shah, 2008; Vieira et al., 2012).

Because the underlying causes of pregnancy-related risks are different for younger and older women, they are likely to follow different treatment patterns. Thus, selection into treatment is potentially driving the heterogeneous results found for capitation contracts. Online appendix figure A4 supports this theory. The figure shows that the number of GP visits for first-time mothers decreases marginally by maternal age but also that the number of specialist visits (all types) increases by maternal age. For 1990, Joergensen (1993) also finds that women older than age 35 had more than twice the chance of having an ultrasound screening (by an OB).

In addition, the literature suggests that risk-related behavior decreases fetal health. As maternal education is a good proxy for risk behavior, I investigate the impact of maternal education on the effects of capitation contracts on infant health (online appendix table A6). I find that mainly the less educated women (with only ten years of schooling) drive the estimates for all women and the younger women, whereas the results for the less educated older women are insignificant and less systematic in any direction. Together, these findings support the theory that underlying differences in the risk factors (biological versus behavioral) generate selection into treatment and thus heterogeneous effects of capitation contracts.

An alternative explanation to the selection of older and younger women into treatment is that better educated women have higher bargaining power. Consequently, the reform would affect the better educated women to a smaller degree, as they are more likely to obtain a referral to a specialist or maintain a high level of GP visits. To some extent, table A6 also supports this theory, because the estimates for women with a college degree are in most cases smaller than those for the women with ten years of schooling.

Thus, as GPs’ personal gain under capitation contracts increases not money but leisure time, these GPs are less likely to deviate from a standard level of care. After the reform, GPs had higher economic incentives to increase the amount of treatment per patient and reduce the number of referrals to specialists (Krasnik et al., 1990). For older women, this change in payment system had little effect; older women with a higher level of bargaining power and a higher level of pregnancy-related biological risks demanded a higher level of specialist treatment. In contrast, patients such as younger and less educated women who are also in need of additional care (but, e.g., did not have the bargaining power), are more likely to benefit from incentivizing GPs.

C. Robustness Checks

Growing public awareness about the benefits of a healthy lifestyle during pregnancy is a likely cause of the general improvement of infant health status. For example, birth weight increased by 130 grams in Denmark from 1980 through 2003 (Sundhedsstyrelsen, 2004a). If this increased awareness has developed differently in the treatment than in the control groups, a possible risk of upward bias exists.

If I assume that the within-group specific trends are constant across time, a DD model similar to model (1) using only data before October 1987 tests for these differential trends. Thus, I apply this robustness check, with births from October 1984 to September 1987 defining the “false” period under FFS contracts. I find that most estimates are closer to 0 than in table 4 and that no outcomes are significant. However, for log birth weight and very preterm birth, I find a marginal significant effect for prime-age mothers (online appendix table A7). Thus results for this group can be questioned.

V. Discussion and Conclusion

This paper exploits a change in a GP remuneration system on infants’ health. I find that only the firstborn infants of younger women, exposed in the womb to the care of GPs on capitation contracts, have significantly worse health outcomes than infants of older women, exposed in the womb to the care of GPs on partly FFS contracts.

Differential treatment patterns potentially drive this heterogeneous age effect for two reasons. First, advanced maternal age directly increases the probability of poor infant outcomes, whereas a higher level of risk behavior

26 For older women, pregnancy-related risks mentioned in the literature include gestational diabetes, high blood pressure, genetic abnormalities, preterm births, and to some extent low birth weight. For younger women, one biological concern for poor infant outcomes is that maternal immaturity may reduce placental functions (Hayward et al., 2011).

27 The figure uses data from 1991 and 1992, as these years are the first where such data are available. I present the total number of specialist visits, as the existing data do not allow me to graph number of visits by type of specialist.
increases the probability of poor infant outcomes for the younger women. Second, as older women are better educated, they have better bargaining power to demand better care or specialist treatments. Additional estimations support this potential explanation of differential treatment patterns.

As no data exist on either the amount or the level of services before the reform, calculating the cost-effectiveness of changing the remuneration system is difficult. Therefore, to contextualize my findings, I first compare them to the well-known effects of maternal schooling on birth weight. Second, as other studies show long-run effects of birth weight, I relate my findings to this literature to frame a possible spillover to long-run effects.

A branch of the economic literature investigates infant health as a social return to maternal schooling, and causal effects of maternal schooling on birth weight are found, for example, in the United States (Currie & Moretti, 2003), the United Kingdom (Chevalier & O’Sullivan, 2007), and Denmark (Bingley et al. 2009). On a U.S. sample of births, Currie and Moretti (2003) use college openings to generate exogenous variation in maternal schooling. They find that an additional 0.9 year of schooling decreases the probability of low birth weight by 2%. A similar strategy by Chevalier and O’Sullivan (2007) finds that the effect of one additional year of schooling in the lower part of the education distribution increases birth weight by 2% to 6% in the United Kingdom. While the U.S. and the U.K. samples estimate effects from either end of the education distribution, Bingley et al. (2009) estimate the average treatment effect of maternal schooling using twin pairs in Denmark. They find that one additional year of maternal schooling increases the average birth weight by 0.9%.

Assuming that the birth weight effect of changing the remuneration system to some extent is comparable to the effects of increasing maternal schooling by one year, the birth weight improvement that I find for younger women falls between half the birth-weight increase from one additional year of maternal schooling in the United Kingdom, two-thirds the increase in the United States, and double the increase in Denmark.

Infant health is also a proven proxy for human capital formation (Currie, 2009; Behrman & Rosenzweig, 2004). Therefore, I also relate my findings to the literature on birth weight to frame a possible spillover to long-run effects. On a Norwegian sample of births, Black et al. (2007) find that a 10% increase in birth weight increases the probability of high school completion by 0.9 percentage points, or 1.2%. Imposing the strong assumption that similar effects occur in Denmark, the effect of changing the remuneration system on high school completion, mediated through a 1.8% higher birth weight, therefore translates into an approximately 1.5% higher probability of high school completion. Using a sample of California births, Royer (2009) finds that a 250 gram or 10% increase in birth weight increases years of schooling by 0.04. Thus, if similar birth-weight effects of changing the remuneration system can be found in the United States, this change would increase years of schooling by 0.08 month.

Relating my results to the literature suggests that my findings are also important to public resource decision making on improving long-run outcomes. However, because I investigate the effects only for infants, I provide no evidence that remuneration changes are effective in improving the health of other groups in the population. In addition, infants born in Denmark during the 1980s may have been particularly responsive to improved primary care. Gaining a broader understanding of the ultimate consequences of incentivizing physicians requires future research into exposures over other periods of the life cycle.

28 All effects reported from previous studies are significant.

REFERENCES


**Sundhedstyrelsen,** *Retningslinjer for Svangreskabshygjiegje og foedselsjaelp* (Copenhagen: Sundhedstyrelsen 1985).


