THE DANISH
NATIONAL CENTRE
FOR SOCIAL RESEARCH

Vibeke Myrup Jensen

13:2009 WORKING PAPER

STARVE THE DOCTOR – STARVE THE BABY?

- Incentivizing Physicians Improves Quality of Prenatal Care

RESEARCH DEPARTMENT OF CHILDREN AND FAMILY

STARVE THE DOCTOR – STARVE THE BABY?

- Incentivizing Physicians Improves Quality of Prenatal Care

Vibeke Myrup Jensen

Children and Family
Working Paper 13:2009

Starve the Doctor – Starve the Baby?
– Incentivizing Physicians Improves Quality of Prenatal Care*

Vibeke Myrup Jensen

The Danish National Centre for Social Research, Copenhagen

September 2009

Keywords: Payment reform, physician-induced demand, child health

Physician-induced demand is a central cost-efficiency issue in public health care programs but little is known about effects of remuneration systems on the ultimate outcome – patient health. Exploiting a unique combination of policy-induced variation and administrative records from Denmark, I investigate the impact of GP payment contracts on infants' health. In a difference-in-differences framework, I find that first-born infants exposed in the womb to the care of GPs with capitation contracts have poorer infant health outcomes relative to infants exposed to fee-for-service contracts. Further I find the estimated effects primarily are driven by younger mothers. (JEL I11, I18, I19, J13)

*Acknowledgements: I thank Anders Björklund, Erik Plug, Nabanita Datta Gupta, Dayanand Manoli, Kathleen Mullen, Paul Bingley, Allan Krasnik, Tor Eriksson, Sandra E. Black, Paul Devereux, participants at IHEA World Conference, Beijing 2009, the workshop in Applied Microeconomics and Programme Evaluation at SFI 2009, the SFI health working group, and ASB economics seminar for useful comments.

Contact information: vmj@sfi.dk; vimj@asb.dk.

1. INTRODUCTION

Physician-induced demand, whereby physicians deviate from providing the optimal health care to pursue personal gain, is a central cost-efficiency issue in public health care programs, and is at the heart of concerns about publicly provided health care such as the Medicare program in the United States. Several studies investigate the effects of incentivizing physicians and find this to increase the number of services provided. However, no studies investigate how incentivizing physicians affects the ultimate outcome, the health of their patients. This paper fills a gap in this literature by estimating the effect of policy-induced changes in physician remuneration system on infant health.

In order to identify the effects of remuneration changes on newborns I take advantage of an influential natural experiment subjected to general practitioners (GPs) in Copenhagen city. In 1987, GPs in Copenhagen city moved from pure capitation to a mix of one-third capitation and two-thirds fee-for-service (FFS) contracts, bringing them into line with the existing contracts in the rest of the country. A previous study (Krasnik et al. 1990) estimates the effect of this Copenhagen reform on practice behavior and find this to increase the self-provision of services by reduced referral rates (to specialists and hospitals), and increased number of diagnostic and curative services. While Krasnik et al. (1990) provide evidence that the reform had an effect on doctor behavior, the key question is, how did this ultimately affect the patients?

Specifically, I estimate a difference-in-differences model (DD). The ‘treatment’ of interest is infants exposed in the womb to the care of physicians with capitation contracts relative to partial FFS contracts, and the outcomes are measured at birth.

The Danish context is ideally suited to this task. First, selection into treatment is not an issue, because of universal mandatory health insurance coverage (financed through general taxation), and a GP listing system based on residential address. Second, recall-bias is non-existing in that I use, which also allow me to link hospital records directly to parental background information and information about the municipality of residence.

While the change in remuneration system provides an administrative source of variability in treatment exposure among mothers-to-be, I take several steps to establish a counterfactual group. First, I define the geographically adjacent municipalities surrounding Copenhagen City – namely Copenhagen County – as the control group. Given short geographical distances between treatment and control areas and high-level of infrastructure, similarity in job-opportunities between treatment and control areas are high. Second, I add a detailed list of parental- and municipality-level controls. Third, I apply restrictions to the model by estimating this as a municipality-fixed effects model.

Thus, to the extent that the only important difference in county policies is the remuneration system, I find that the health of firstborn infants born to mothers seeing GPs working under capitation contracts during their pregnancy have significantly worse health outcomes. For example, birth weight is 1.5% lower, the probability of preterm birth is 39.4% higher, fetal growth is 1.4% lower, and the probability of receiving the best heath condition on the one and five minutes APGAR scores are 20% and 25% lower, respectively. Furthermore, two triple-difference models show that these results are robust to the potential bias of differential trends and group-specific trends. In addition, I find heterogeneous responses to prenatal care, with a strong effect for mothers under 27 years of age and nothing for mothers above 27. This result suggests that FFS contracts incentivize GPs to additional deviate from a standard quality of prenatal care and patients in need of this additional care such as young mothers' production of infant health benefits from this.

Calculating the cost-efficiency of incentivizing GPs is difficult. Using results from a previous study of this reform Krasnik et al. (1990), physician-induced demand in Denmark estimates to 37% in medical services. Imposing the assumption that the total level of inducement in Denmark translates directly into inducement in prenatal care, a 37% inducement could increase birth weight by 1.5%.

However, infant health is also important for later child development. Research in both the health sciences (e.g. Barker et al. 1995; Power, Bartley & Smith 1996; Anand et al. 2008) and the social sciences (Behrman & Rosenzweig 2004; Almond, Chay, & Lee 2005; Currie 2009) has established the persistence of early health measures into labor market attainment, adult health, and the intergenerational transmission of health. The results are therefore also important inputs into public resource decision making to improve child development. Relating my findings to existing studies of long-run effects of infant health of Black, Devereux & Salvanes (2007) and Royer (2009) indicate that capitation contracts relative to FFS contracts translate into a lower probability of high school completion by 0.07-0.15 percentage points and a 0.26% lower birth weight for the next generation. Further, comparing my findings to the effects of one additional year of maternal schooling, the birth weight improvement of incentivizing GPs is between half the birth weight increase (Chevalier & O'Sullivan 2007), two thirds the birth weight increase (Currie & Moretti 2003), and one and three quarters the birth weight increase as one additional year of maternal schooling (Bingley, Christensen & Jensen 2009).

The remainder of the paper proceeds as follows. Section 2 discusses the related literature on the effect of physicians’ remuneration contracts on medical practice. Section 3 outlines Danish GPs’ remuneration and its reform. Section 4 describes the data and section 5 presents the estimated model. Section 6 presents the results, and section 7 includes some robustness checks. Section 8 concludes, and results are discussed in section 9.

2. PREVIOUS LITTERATURE

GP behaviour is a central issue in health production because of asymmetric information between GP and patient. The existing literature covers the relationship between financial incentives and medical practice as a classical income-leisure model; GPs value their own leisure activities against their income, time constraints, and patient objectives (Scott 2000). However, variation in the theoretical model exists. In particular medical ethics enter either as a utility for the delivery of optimal health care (McGuire & Pauly 1991) or, in line with general human capital theory, as a derived satisfaction from professional good practice (Becker 1965; Woodward & Warren-Boulton 1984).

A central theoretical issue first formulated by Evans (1974) and Fuchs (1978) is supplier induced demand. When the GP as a fully informed agent for his or her patients maximizes his or her utility function, he or she is, at the margin, willing to surrender a certain degree of accuracy about the optimal health care to increase personal gain.

Empirical investigations of physician-induced demand have frequently been reviewed (Gosden et al. 2000; Gosden, Pedersen & Torgerson 1999; Scott & Hall 1995; Town et al. 2005). Gosden et al. (2000) investigate the direct effect of changes in payment systems, where the primary search criteria are randomized control trials or controlled before-and-after studies. Four studies meet their criteria, and all conclude that fixed payment-based systems constrain medical care compared to the FFS or bonus paid systems. The review of Town et al. (2005) focuses on preventive care comparing bonus pay systems or FFS systems to fixed payment systems and in total six studies is reviewed. One out of the six studies, a study investigating the effect of bonus pay system on the rate of immunization finds positive effects of bonus performance (Kouides et al. 1998).

As the choice of payment contracts is often opportunistic, a recent study by Devlin & Sarma (2008) estimates the effect of selection into remuneration contracts. They find a positive selection effect on the choice of remuneration. Taking this selection into account, they find that FFS contracts encourage treating more patients.

While the incentive schemes in FFS systems are different from pay-for-performance systems or value-based purchasing systems, little evidence exists also of pay-for-performance systems improving quality of care. E.g. using introduced health care programs in California, Mullen, Frank and Rosenthal (2009) investigate whether the introduction of pay-for-performance systems was effective of improving quality of care. While the programs did increase the medical measures rewarded by the program they find no improvement of the overall quality of care.

Although not targeting GPs but obstetric/gynaecologic physicians, a related paper by Gruber and Owings (1996) investigates how declining fertility in the U.S. from 1970 to 1982 led physicians to increase the rate of caesarean deliveries. They find a strong correlation between within-state fertility decline and within-state increases in caesarean deliveries, a finding that they interpret as evidence of physician-induced demand, given that caesarean deliveries are more highly reimbursed.

Krasnik et al. (1990) estimate the effect of this remuneration change on practice behaviour by comparing GPs in the city of Copenhagen itself and GPs in the greater Copenhagen area (Copenhagen County) before and after the implementation of the new contracts. They find that GPs' changing contracts increase the self-provision of services by reduced referral rates (to specialists and hospitals), increased the rates of diagnostic and curative services, increased the use of laboratory tests, and increased the number of telephone consultations. I go further than the Krasnik analyses of the Copenhagen reform and move the induced demand literature forward by focusing on ultimate health outcomes rather than services immediately provided.

3. BACKGROUND

3.1. Defining Treatment and Control Areas

Following discussions about workload and income differences between GPs in different counties, GP remuneration systems unified throughout the country in 1987. GP income in the municipalities of Copenhagen, Frederiksberg, Taarnby and Dragoer (hereafter called Copenhagen city) fell behind GP income in other parts of the country. Increasing workloads per patient are the main causes of the income differences as waiting times for hospitalization increased, hospitals started discharging patients earlier, and the proportion of elderly patients increased (Krasnik et al. 1990). 1st of October 1987 GPs in Copenhagen city shifted from pure capitation contracts to the elsewhere in Denmark standard mixed capitation and FFS contracts. FFS makes up two thirds of the GP remuneration system and includes special fees for 4 types of consultation; 39 types of medical services; and 37 types of laboratory tests (National Health Security System & Association of General Practitioners 1984).

From a research perspective, the change in remuneration system is highly attractive, given the natural experiment setting where some areas were exposed to remuneration changes and others were not. The adjacent municipalities defining Copenhagen County as used in Krasnik et al. (1990) are chosen as the control area whereas the municipalities defining Copenhagen city are defined as the treatment area. Short geographical distances – approx 25km from capital centre to the outskirts of Copenhagen County – and cheap public transport make commuting easy and diminish differences in GPs listed patients.

Figure 1 shows the gross annual income distribution (in 2008 USD and reflated) of GPs in Copenhagen city and Copenhagen County during 1980-1995[1]. Before 1987 GPs income in Copenhagen city is consistently USD 4000-5000 (7.7%) below the level of GPs in Copenhagen County. In 1987 the income gap narrows and is more similar thereafter.

However, for this type of evaluation to be valid, treatment and control areas need to be similar to one another before time of event and to follow same trends to ensure unbiased inference. The following subsections discuss similarities and changes in health care provision in the two areas whereas next section on data discusses differences in population.

3.2. The Danish Health Care System and GP Contracts

Health care insurance in Denmark is publicly provided and financed through general taxation. Ninety-five per cent of the population is covered[2], which reduces a systematic selection of patients into health insurance coverage. Counties are the local health authorities, but the central government also seeks to increase national homogeneity by indirect regulations of county service provision by way of block grants and national recommendations.

Each GP has a contract with the local health authority and needs to follow national regulations to assure equality in health care access across counties. As in many countries, Danish GPs act as gatekeepers for all types of non-emergency health care which minimises group composition changes due to differential trends in specialist treatments (Scott 2000; Vallgaarda et al. 2001).

Binding agreements between the National Health Security System (NHSS) and the Association of General Practitioners define a detailed set of regulations licensing medical practice for all GPs in the country both before and after the remuneration system changed in Copenhagen city. Regulations include office hours, patient treatment, supply-side controls, and GP:patient ratios. To keep GPs from choosing some patients over others, patients

---

[1] The figure is adjusted for municipality-level differences in GP age, GP tenure, and municipality-, county-, and church-tax.

[2] Health insurance includes all medical expenses except pharmaceuticals and dentistry. The alternative health care option is a 100% payment out-of-pocket as a quid pro quo for exemption from the GP gatekeeper and listing system.

choose their GPs from a list according to residential area. Within the limits of patients per GP and the practice district, each patient chooses a GP from publicly available information about practice size and the GP's gender and age.

The overall relationship between a patient and the GP is characterized by continuity of care. Only once a year the patient can change GP, and children are automatically assigned to their mother's GP. Anticipation behaviour by patients due to remuneration changes is therefore less likely to happen which is further supported by the low level of public awareness of method of GP payment at the time.

The detailed national regulations minimize differences between the GP practices in Copenhagen city and their counterparts elsewhere in Denmark, but Copenhagen city likely became more attractive to different types of GPs after the reform. As a consequence, the flow of new GPs could change. However, this causes limited implications because all GP education is state-funded and subject to the same guidelines and curriculum standards. Further, as described the speed of changes to the stock is limited by licensing of practices according to GP:patient ratios.

While the stock of GPs can change slowly, different trends in the local tax level increase an immediate risk of shifting incentive levels which may dilute the effect of remuneration changes. Figure A1 in the appendix graphs the tax-level distribution across time, pooling county, municipality, and church tax for Copenhagen city area and Copenhagen county area in 1980-1995. For both areas there was a slightly positive trend in the tax-levels, but the local tax-levels were very similar throughout the period. Further, given that all GPs are in the same income-group, state tax has no influence on differences in the incentive schemes, which minimizes a potential bias from tax-related differences in the incentive levels.

3.3. Prenatal Health Care

Preventive care plays a central role in general practice, and regular prenatal consultations have been a standard service since 1945 (Vedsted et al. 2005). A pregnant woman is by law guaranteed free of out of pocket-expenses midwifery, three doctor consultations and five midwife consultations. Midwife regulations and prenatal care recommendations are nationally regulated and there were no major changes affecting either the treatment or the control group throughout the period that could have offset differences in prenatal care (Ministry of the Interior and Health 1985; Ministry of the Interior and Health 1994). However, unlike many other European countries there exist no guidelines in terms of the number of obstetric ultrasound screenings (OUS) per pregnancy in Denmark (Joergensen 2002).

OUS is an important part of modern prenatal care and the use thereof changed rapidly from the late 1970s until today (Joergensen 2002; Joergensen 2003). Therefore, a potential bias exists if the implementations of OUS in prenatal care or changes in OUS equipment happen faster in one area compared to the other.

In 1990 the provided number of OUS was very similar for the treatment and control areas. 94.6% of all women had at least one OUS in Copenhagen city whereas the number was 93.5% in Copenhagen county. Further, the average number of OUS per pregnancy was 2.2 in Copenhagen city and 1.7 in Copenhagen county (Joergensen 2003). Although not providing direct evidence of the development of OUS treatment, these similarities in the levels of OUS between the treatment and the control areas are comforting.

Assuming a relationship between the demand for OUS and the supply of OUS, I can use the development in hospital births as a proxy for the development in the supply of OUS, because almost all OUS are conducted at the hospital (Joergensen 2003). Figure A2 in the appendix shows the distribution of hospital births at the six hospitals with a birth clinic in the treatment and control areas. As indicated by the figure, the average number of hospital births is similar in the treatment and control areas and the changes over time follow similar trends.

4. DATA

4.1. Data

The dataset is high-quality medical records from a 15% random sample of firstborn infants born 1982-1992. I focus solely on firstborn infants for a clearer causal interpretation. Because first-time parents are more sensitive to prenatal care, and I eliminate dynamic intra-household investment behaviour where experiences from the first birth affect choices linked to the second birth (Del Bono, Ermisch & Francesconi 2008). Although, the restriction to firstborns also implies a limitation, because generalizing the results to non-firstborn infants is questionable.

The birth records contain information on exact birth date, a wide range of infant health measures, and information on the mother such as age and previous stillbirths. Using unique social security numbers, I match infant medical records with parental background information from the Integrated Database for Labor Market Research. A database created from several administrative registers covering the Danish population from 1980 and maintained by Statistics Denmark. My dataset contains information about parental educational attainment, earnings, occupation and demographics such as marital status, country of birth and parental birth date.

The outcome of interest is infants' health status. I use information on birth weight, gestation, and APGAR scores. As birth weight measures I use actual birth weight, which is observed in ten gram intervals, and a dummy for low birth weight using the WHO threshold of 2500 grams (World Health Organisation & The United Nations Children's Fund 2004). I use gestation measured in weeks converted into a binary variable for preterm birth, using the WHO definition of less than 37 full weeks (World Health Organisation & The United Nations Children's Fund 2004). To account for complex interaction effects between birth weight and gestation, I use the commonly used measure of foetal growth defined as birth weight in grams divided by weeks of gestation. The APGAR score summarizes an infant's health condition by measuring health on five criteria: Appearance, Pulse, Grimace, Activity, and Respiration – each with the possible values of zero, one or two. The sum of these five criteria defines infant health status, and all infants are evaluated one and five minutes after birth. I convert these two measures into dummy variables – with unity defining the best possible health condition.

4.2. Summary Statistics

Table 1 provides sample mean and standard deviations for the outcomes of interest and the extensive list of covariates during the pre- and post-reform periods for treatment and control groups, separately. In general, the treated and control groups are very similar in terms of background information, but more important the changes from one period to the next happen at the same rate. The largest difference is found in maternal age at child birth. In the treatment group mothers' age increases by 0.68 year from the pre- to the post-reform period whereas mothers' age increases by 0.91 years in the control group.

Distributions of birth weight, gestation, foetal growth, and five minutes APGAR score are graphed in Figure 2. These are split by maternal age groups: above and below the average age of first child birth (26 years old) – hence older and younger mothers. In particular for younger mothers in the treatment group, the outcome distribution shifts towards better outcomes in the post-reform period compared to the pre-reform period. The picture is less clear for the older mothers. Differences in outcomes between younger and older mothers are more evident when graphed over time. Figure 3 illustrates this point by OLS estimations of infant outcome on birth year (1982-1992) for the treated and control groups, and younger and older mothers, respectively. I graph the means for birth weight, foetal growth, one minute and five minutes APGAR score, and the OLS estimations include the list covariates in Table 1. The vertical line in 1987 illustrates the year of remuneration change. While the outcome distribution across birth years is almost identical for treated and control groups when the mothers are older, a gap between treated and control groups before 1987 is evident for the sample of younger mothers.

5. CONCEPTUAL FRAMEWORK

In this section, to measure the effect of changing remuneration on infant health I formulate the difference-in-differences (DD) model specification.

5.1. Identification Strategy

The key estimation for infant *i* in municipality *m* at time *t* is the following:

$$Health_{imt} = \beta_{0m} + \beta_1(Cph\_City_m * policy_t) + \beta_3 year_t + X_{imt}\beta_4 + \varepsilon_{imt}, \qquad (1)$$

where the dependent variable $Health_{imt}$ is different measures of infant health; Cph_City_m is a binary equal to one in the treated municipalities in the Copenhagen city area and equal to zero in the municipalities of Copenhagen County. This variable is interacted with a '$policy_t$' variable, a binary variable where one defines the five-year period where GPs in Copenhagen city were on capitation contracts , and zero defines the five-year period after mixed contracts of capitation and FFS were implemented in Copenhagen; $year_t$ defines a full set of birth year dummies to control for general changes over time; X_{imt} defines a vector of parental and family background variables to control for potential heterogeneity in time trends between treated and control groups.

Each treatment and control area consists of a number of municipalities, therefore the longitudinal data allows estimation of a municipality fixed effect model. This is important to eliminate case-load differences between treatment and control groups. Thus β_{0m} corresponds to the municipality level intercept and the binary indicator Cph_City_m defining the treatment and control municipalities is dropped from the model.

I use OLS to estimate equation (1). I correct for an arbitrary autocorrelation in the error term by computing White-robust standard errors. A method proven to provide consistent estimations for datasets with a large number of N and few periods (Bertrand, Duflo & Mullainathan 2004).[3]

The reduced-form effect of capitation compared to mixed capitation and FFS contracts is identified by β_1, and this DD estimator is grounded on two main assumptions; first, conditional on the list of covariates, the composition of the group does not change over time. Second, there exists a counterfactual relation between treatment and control groups (Blundell & Costa Dias 2000). Section 3 discussed both assumptions extensively and the model, estimated with municipality fixed effects, includes an extensive list of covariates. A potential bias still exists if differences by some unknown factor in treatment and control area makes GPs or patients react differently to common time-specific changes. I therefore test this assumption in section 7 on robustness checks.

5.2. How Can GP Contracts Affect Infant Health?

For financial incentives to affect infant health, relevant GP services improving prenatal care need to be included in FFS rates. There are mainly two channels in which GPs can improve the health of the unborn child. These are: additional tests and additional consultations. An increased number of tests and curative services are important to prevent and cure infections unhealthy for foetus growth and to prevent premature births. The national recommendations are urine and blood tests that include tests for infections, pre-eclampsia, hepatitis, diabetes, hereditary diseases, etc. (Ministry of the Interior and Health 1985).

Because GPs and midwives are obligated to provide counselling in terms of appropriate lifestyle during pregnancy, additional face-to-face or telephone consultations are helpful in assuring sufficient nutrition for the foetus. Standard recommendations include diet, physical activity, restricting consumption of alcohol and tobacco, and mentally preparation of the mother and father for their new parenting role.[4] Finally, GPs and midwifes also have

[3] Further, including clusters in the fixed effect model for group dependence in the error term does not affect the results negatively. Therefore the model presented is estimated without any clustering.
[4] While risk behaviour like smoking, drug and alcohol use is well known for its unhealthy effects on foetus growth, an U.S. study suggests that depressions during pregnancy increases the risk of premature births (Su-

to report to the social services if the mother-to-be seems in any kind of social or financial difficulty that in any way affects the health of the unborn child (Ministry of the Interior and Health 1985).

Prenatal medical services also need to be included in FFS rates. Measuring GP's prenatal care activities directly in the period of investigation is impossible. Instead I use GP prescription data from 1997-1999 to motivate that the relevant services for prenatal care are included in FFS rates. Table 2 lists the seven most commonly used FFS items during pregnancy and the distribution of these services on pregnancy vs. non-pregnancy related consultations. Further, the fee-per-item as listed in the 1984 agreement between NHSS and the Association of GPs and a calculated average fee per pregnancy are included in the table.[5]

During a pregnancy period the average number of consultations are 6.4 face-to-face and 3.4 by telephone, whereas the average number of consultations for a woman not carrying a child are 1.8 face-to-face visits and 1.2 telephone consultations. Although the primary content of a GP visit is unknown for both the groups of pregnant and non-pregnant women, the difference between the high rate of GP visits during pregnancy and the low rate for non-pregnant women suggests that consultations during pregnancy are pregnancy-related. The number of blood and urine tests is also higher during pregnancy. Calculating the average cost of a pregnancy, the blood and urine samples include the sampling fee whereas the blood and urine laboratory tests include the testing fees. Therefore, the testing fees can be underestimated if laboratory tests are only reported under laboratory test but credited under both sampling and testing. The fees for an average pregnancy total USD 66.2 and USD 14.3 for the same type of consultations for non-pregnant women, which indicates a higher GP income related to a pregnancy. In comparison, the remuneration that GPs received from the annual capitation per patient fee was USD 26.9.

---

giura-Ogasawarna et al. 2002). Recent descriptive statistics from the National Board of Health (2007b) support this tendency.

[5] These medical care statistics are based on a 5% random sample of women carrying their firstborn child in 1997-1999 and a 5% random sample of similarly aged women not carrying a child.

6. RESULTS

I use a DD model to estimate the effect of remuneration on the quality of prenatal care and report estimates for the full sample, and separately for younger and older mothers. Younger mothers-to-be have been shown to need extra prenatal care, because the pregnancy is less likely to be planned, and thus involving delayed initiation of prenatal care. Furthermore, to remain slim, younger women are more likely to resist recommended dietary and calorie intakes, test social boundaries through increased risk behaviour such as drinking and smoking (for a review of the literature, see Ohlsson & Shah 2008). Thus, patients such as younger mothers-to-be, with treatment needs beyond the standard provided care, are at risk of receiving insufficient health care if GPs under capitation contracts have little financial incentive to increase medical services as suggested by Scott (2000).

6.1. The Effect of Remuneration on Infant Health

Estimates for the full sample, younger and older mothers on infant outcomes with different sets of controls, appear in Table 3. Columns 1-2 report the full sample for two sets of controls: no controls, and parental background and GP characteristics (age, age squared and gender). Columns 3-4 report outcomes when the mother is under 27. Columns 5-6 report infant health outcomes when the mother is 27 or above. The results support the hypothesis that capitation contracts decrease incentives to improve the quality of health care.

Full sample.—Estimating the effect of the remuneration system on infant health using the full sample and a model including no controls, the results show that birth weight is 53.7 grams lower (1.6%) in the capitation system. The estimate is -50.8 grams equivalent to 1.5% lower average birth weight when the full list of parental background, family characteristics and GP information are included. As with the effect of remuneration on birth weight, including covariates in the model has little impact on the other outcomes. This fact indicates that the municipality fixed effects model picks up most differences between treated and control groups. The effect on low birth weight is both small and imprecisely estimated. Taken together with the previous average effect on birth weight, this result indicates that the effects of financial incentives investigated are not sufficient to target problem low birth weight (of clinical significance).

The probability of pre-term birth is 2.1 percentage point higher equivalent to 39.4% under capitation contracts. Foetal growth rate is 1.2 gram per week (1.4%) slower. Finally APGAR scores, measuring vital status both one and five minutes after birth, show a lower probability of optimal health status (score=10) by 10 percentage points equivalent to 20.6% and 25.9%, respectively.

Mothers under 27.—Columns 4-6 report estimates for younger mothers. Given the reduction in sample size, the estimates reported here are more sensitive to inclusion of the covariates. Infants born under the capitation system have on average an 89.3 gram lower birth weight (2.7%) compared to infants born under the mixed system of capitation and FFS, when no controls are included. The effect is -80.6 grams (-2.4%) when the full set of parental background and GP information is included. For younger mothers there are also no effects on the probability of low birth weight; the probability of preterm birth is 2.6 percentage points (56%) higher, and foetal growth is 2.0 grams per week (2.3%) slower. Furthermore, the probability of receiving the best possible health condition on the one-minute APGAR score is -9.7 percentage points (21.4%).

Mothers 27 or above.— Columns 7-9 report outcomes for older mothers. With two exceptions I find little effect of the remuneration system on infant health for this group of mothers. For example, the effect of capitation contracts on birth weight is -4.2 grams, indicating that changes in remuneration had little impact for this group. The exceptions are APGAR scores. At one minute score the probability of receiving the best possible health condition is 9.4 percentage point lower under the capitation system (20.3%). The effect after the five minutes score is a lower probability of 3.4 percentage points (40.1%), although, the latter estimate is only significant at a 10 percent level.

7. ROBUSTNESS CHECKS

As noted in section 5, the DD estimator is based on two assumptions: first, changing remuneration provides exogenous variation in prenatal care, and second, the existence of a counterfactual relationship between treated and control groups. Following Blundell & Costa Dias (2000), I can decompose the error term from the DD municipality fixed-effects model into:

$$\varepsilon_{imt} = \phi_m + \theta_t + \mu_{imt}\,, \tag{2}$$

where ϕ_m is an municipal-specific effect constant over time, θ_t is a time-specific change common to all pregnant women, and μ_{imt} is a temporary individual- and municipality-specific effect. As stated in section 4, anticipation effects – i.e. the effect of the individual- and municipality-specific error term μ_{imt} – is unlikely to affect the outcome. However, the second assumption defining the counterfactual relationship between treated and control groups can be violated given groups are defined by geographical areas. The problem can be written as the following:

$$E(\varepsilon_{mt} \mid treated_m) = E(\phi_m \mid treated_m) + k_g\theta_t \tag{3}$$

where k_g defines the group-specific effects of time-specific changes θ_t across the two groups. Thus the DD estimator is defined as:

$$E(\hat{\beta}_{DD}) = \beta + (k_T - k_C)(\theta_{t1} – \theta_{t0}) \tag{4}$$

where subscripts T and C refer to the treated and control groups. Then $\hat{\beta}$ is the true effect of remuneration changes when the control group is counterfactual to the treated, i.e. $k_T = k_C$, or the time specific effects are similar across groups $(\theta_{t1} = \theta_{t0})$.

I check this assumption in two ways. I adjust for time-specific trends affecting treated and control groups differently by estimating two triple difference models (DDD). First, by including two artificial treatment periods, and second, by including artificial treated and control groups from other parts of the country. Given both models approach the problem from different angles and estimates are similar in magnitude, I argue that the DD model generates robust effects of remuneration changes.

7.1. Adjustment by Including Additional Time Periods

Growing public awareness about the benefits of living a healthy life style during pregnancy is a likely cause for generally improving infant health status. One indication of this is that birth weight increased constantly during the period. Therefore a possible risk of upward bias in the estimations is evident, potentially generated by a different effect of the increased awareness of healthy lifestyle benefits during pregnancy between treatment and control groups. The DDD model will adjust for these differential time trends assuming that the within-group specific trend is constant across time.

In the context of a health care reform, choosing a second pair of pre-reform or post-reform periods to check for differential time trends is not obvious. Given that the timing of this reform lies closer to the beginning of available data choosing an additional post-reform comparison is only feasible. Furthermore, including an additional ten-year period violates the issue of using recent periods for time trend adjustment. Therefore, I narrow the original birth year window to three years before and after the policy implementation in Copenhagen city (1985-1990) and use the following sequence of birth years (1991-1996) as 'artificial' before- and after-reform periods to adjust for differences in time trends between Copenhagen city and Copenhagen county.[6] The maintained hypothesis is that the differences in time trends between treatment and control groups are persistent.

---

[6] In this case narrowing the time-window is necessary, because the probability of other things changing and thereby adding noise to the estimation increases by each additional year one moves away from implementation of the policy. However, the choice of a 3-year window is also a trade-off between reducing sampling variability and sufficient sample size to conduct a meaningfull analysis.

The DDD model then takes the following form:

$$Health_{imt} = \beta_{0m} + \beta_1(Cph\_City_m * policy_t) + \beta_2(Cph\_City_m * treat\_period_t) +$$
$$\beta_3(policy_t * treat\_period_t) + \beta_4(Cph\_City_m * treat\_period_t * policy_t) + \quad (5)$$
$$\beta_5 treat\_period_t + \beta_6 year_t + X_{imt}\beta_7 + \varepsilon_{imt}$$

The dependent variable $Health_{imt}$ defines infant health outcome. As in model (1), Cph_City_m defines a dummy variable equal to one in municipalities in the Copenhagen city area and zero in the county area; X_{imt} is a vector of covariates; and ε_{imt} is the error term. The dummy defining the treatment group Cph_City_m is interacted with two other variables defining treatment periods $policy_t$ to identify the effect of capitation contracts and $treat\_period_t$ for common time trend adjustment. $policy_t$ defines a variable equal to one in the 1985-1987 period where remuneration consisted of capitation contracts in Copenhagen city and one in the new 'artificial' treatment period 1991-1993, whereas the variable is zero in both post-treatment periods 1988-1990 and 1994-1996. $treat\_period_t$ equals one in the original periods under investigation 1985-1990 and zero in the time trend adjusting periods 1991-1996. $year_t$ defines a full set of year dummies. As in model (1), I include municipality fixed effects which make β_{0m} municipality-specific constants and the main effect of Cph_City_m defining treatment group is dropped from the model. The coefficient of interest is captured by β_4, the interaction term defining the time-trend adjusted effect of capitation contracts compared to mixed capitation and FFS contracts.

Given that I restrict the birth year window to three years before and after remuneration changes, table 4 re-estimates the specifications of Table 3 using the reduced sample. In general, the effects are similar in magnitude but imprecisely estimated because of a 47% sample size reduction (from approx. 9200 infants to 4900 infants).

Table 5 reports estimates of the DDD model. As in model (1), I report separate effects for the full sample of mothers, younger and older mothers. In general, the effects of birth weight and foetal growth are larger compared to those of Table 3 whereas the effects on APGAR scores are reduced.

Full sample.—Using the full sample, the trend-adjusted effect of capitation is -2.3% in birth weight, and a higher probability of preterm birth by 3.9 percentage points. Only the effect on preterm birth is significant at a 5% level, whereas the effect on birth weight is significant at a 10% level. Other outcomes are not significant.

Mothers under 27.—Using the sample of younger mothers, the effect of capitation is 3.6% lower birth weight, and 0.2% slower foetal growth, whereas the effect on low birth weight and the APGAR scores are not significant. Point estimates of APGAR score effects are reduced by half.

Mothers 27 or above.—The main table (Table 3) showed little effect of capitation contracts on infant health for older mothers. A similar trend is evident when I use the DDD model. Only the effect of capitation on preterm birth is marginally significant where capitation lowers the probability of preterm birth by 0.05 percentage points. However, as for the group of mothers below the age of 26 I find stronger effects for birth weight, preterm birth, and foetal growth whereas the impact on APGAR scores is reduced by half.

In summary, I find greater spread in outcomes compared to estimates using the DD model. The effect on birth weight, gestation, and foetal growth showed larger, but less precisely estimated, effects. This suggests that for these outcomes the DD model provides lower bound estimates. For the APGAR scores, I find the effects to be reduced by half.

One explanation for this diverse effect for different outcomes is increased capacity in the maternity wards as also indicated by figure A2 in the appendix. From 1982 to 1992, the number of births delivered at a hospital increased nationally from 97.2% to 99%, and during the same period maternity wards were integrated inside fewer hospitals, each with greater capacity. The rate of still births, perinatal deaths and low APGAR score (below 7), which are indicators of quality of delivery, were constant during this period (The National Board of Health 2007a). However, given the APGAR scores, unlike birth weight, are sensitive to complications during birth, the preventions of, and faster reactions to, delivery complications at better equipped hospitals could on the margin have improved the probability of highest APGAR score without having a direct impact on low scores.

Furthermore it can be questioned, whether the inclusion of an additional period to adjust for differential time trends make up for the loss of precision due to reducing the sample size. To address the former issue I, as an alternative strategy, follow the ideas of (Dranove & Wehner 1994) and investigate if there is a treatment effect in two post-treatment periods (1987-1991 and 1992-1996) where no effects are supposed to be found. In general, the outcomes pass this falsification test and therefore these effects are not reported.

7.2 Differences between Inner City and County Residential Area

The previous section analyzed differential time trends between groups. This section focuses on differences between treated and control groups by including additional ‘placebo’ treatment and control groups. Copenhagen city and Copenhagen county may have responded differently to other national policy changes coincidental to the timing of the reform. For example, hospital capacity could have changed more rapidly in the county area during the treatment period given a higher probability of an already high rate of hospital births in the city area because of a shorter distance to a hospital. Estimating another DDD model including placebo treatment and control groups for the inner city and the surrounding areas takes into account such differences.

I use the second largest city (Aarhus) and the greater Aarhus area as adjusting treatment and control groups. For this strategy to be valid, the maintained hypothesis is that the relative difference between Aarhus and the greater Aarhus area is similar to the relative difference between Copenhagen city and county, i.e. that Aarhus and the greater Aarhus area match the term $(k_T - k_C)(\theta_{t1} - \theta_{t0})$ from model (2). If this assumption is correct, the adjusting DDD estimator yields a consistent estimate of changing remuneration for individual *i* in municipality *m* at time *t*:

$$Health_{imt} = \beta_{0m} + \beta_1(City_m * policy_t) + \beta_2(treated_m * policy_t) + \beta_3(City_m * policy_t * treated_m) + \beta_7 year_t + X_{imt}\beta_4 + \varepsilon_{imt} \qquad (6)$$

As in model (1) and (5) $Health_{imt}$ the infant health outcome; $year_t$, a full set of year dummies; X_{imt}, a vector of individual-specific covariates; and ε_{imt}, the error term. $policy_t$ defines a dummy equal to one for infants conceived and born between 1982-1987 and zero in the post treatment period 1988-1992. This variable is interacted with $City_m$. A variable which is equal to one in the city areas (Copenhagen and Aarhus cities) and zero in Copenhagen county and greater Aarhus area. $treated_m$ defines the treated groups under investigation and therefore equals to one in Copenhagen and Copenhagen county and zero in Aarhus and greater Aarhus area. The policy impact is identified through the interaction term $City_m * policy_t * treated_m$ that adjusts treated (Copenhagen city) for time and inner city specific time effects. Given the model is estimated as a municipality fixed effect model, then β_{0m} defines a municipality-specific constant term and the terms defining treatment groups $City_m$ and $treated_m$ cancel out.

Figure 4 illustrates the geographical areas of Copenhagen city, Copenhagen county, Aarhus and the greater Aarhus area by mapping the fertility ratio for these groups. While Aarhus and Copenhagen cities show similar fertility profiles, the greater Aarhus area has a higher fertility ratio compared to Copenhagen county. Further, it is evident that the geographical areas are larger because of a lower population density in Aarhus and surrounding areas. Given geographically adjacent areas was one of the reasons for choosing Copenhagen county as a control for Copenhagen city, I define the greater Aarhus area as Aarhus county with exception of the most distant municipalities, further than 50 km away from Aarhus city centre. In total, the population density in Aarhus surrounding area is one fifth of the population density in Aarhus city whereas the population density in Copenhagen county is one third of the population density in Copenhagen city area.

The distribution of outcomes across birth years for the four groups of treated and control groups is further graphed in Figure 5. As in Figure 3 the vertical line in 1987 defines remuneration changes. The figure shows that the distributions of outcomes follow similar trends for the different groups.[7]

Full sample.— Table 6 shows estimates of the DDD model including additional treatment and control groups. Using the full sample, I find that the inner city adjusted effect of capitation on birth weight and foetal growth to be similar compared to the effect found in Table 3, however, they are imprecisely estimated. Unlike the time trend adjustment, the APGAR scores are also similar in size, but only significant using the one-minute score. This suggests that coincident policy changes such as the centralization of maternity wards had a similar impact on treatment and control groups. The effect of capitation on low birth weight remains small and is not significant, and the effect on preterm birth is reduced by half but imprecisely estimated.

Mothers under 27.— Reducing the sample to young mothers, the effects of capitation are similar in magnitude except for birth weight and foetal growth where the effect is larger. This suggests that the main DD model provides lower bound estimates for these outcomes. The effect on birth weight is -3.0% and foetal growth rate is 2.4% slower whereas the effects on low birth weight, preterm birth, the AGPAR scores are similar in magnitude.

*Mothers 27 or above.—*As in Table 3, most effects of capitation are small and imprecisely estimated. The exception is the probability of low birth weight where this probability is 5 percentage points higher under capitation contracts.

In sum, I find similar estimates when adjusting the model for common city and surrounding area trends. Although, these estimates are less precisely estimated.

[7]Given periods are defined by date of conception and date of birth, there are only a few observations in late 1988.

8. CONCLUSION

The health economics literature has moved from establishing the existence of physician-induced demand to quantifying its extent in terms of health care services provided. In this paper I move the literature forward by measuring the net effect of a GP remuneration system reform on the ultimate outcome of interest, the health of the patients.

In 1987, GPs in Copenhagen city were shifted from pure capitation to mixed capitation and FFS contracts, bringing their remuneration contracts into line with other geographical areas in Denmark. Applying a difference-in-differences model, I investigate the effect of capitation contracts on the health of firstborn infants of mothers seeing GPs working under different contractual arrangements during their pregnancies. Furthermore, to allow for differential time trends and group-specific trends, I apply two robustness checks using DDD models. Both robustness checks yield estimates of similar magnitude, but with less precision.

I find that firstborn infants born under capitation contracts had significantly worse health outcomes. This average effect on infant health is generated by the effect of younger mothers (under 27) whereas the effects for the older mothers (age 27 and above) are both small in impact and imprecisely estimated. Taken together with earlier findings from Denmark (Krasnik et al. 1990), the results therefore suggest that the standardized quality of prenatal care likely to be produced under capitation contracts could be improved for patients with additional health care needs. Because GPs under capitation contracts have little incentive to deviate from a standard quality of prenatal care, patients in need of additional care such as younger mothers benefited from remuneration changes whereas infants born of older mothers had no such benefit.

A question that this study cannot answer is whether remuneration changes can be effective in improving the health of other groups of the populations. Infants (born in Denmark during the 1980s) may be particularly responsive to improved primary care. Gaining a broader understanding of the ultimate consequences of incentivising physicians requires considering exposures over other periods in the life cycle. This promises to be a fruitful topic for future research.

9. DISCUSSION

Drawing a firm conclusion on the direct cost effectiveness of changing remuneration is difficult, given that the inducement rate and exact expenses for prenatal care in the pre- and post-reform periods are unknown. Instead, I relate my findings to the existing literature estimating long-run effects of birth weight to frame a possible spill over of the changing physician remuneration system; compare the birth weight estimate to the net inducement rate in Copenhagen; and relate the effect of remuneration on birth weight to papers investigating the effect of maternal schooling on birth weight.

Extrapolating Long-Run Effect of Remuneration System.—In other contexts, infant health is viewed as an input, i.e. as a proxy for early human capital formation that affects the accumulation of human capital later in life (Behrman & Rosenzweig 2004; Black, Devereux, & Salvanes 2007; Oreopoulos, Page & Stevens 2006; Royer 2009). Assuming that infant health effects translate directly into the long-run outcomes found in other studies, extrapolating long-run effects will help understand the total social return to remuneration changes. Using a Norwegian sample of births born between 1967 and 1981, Black, Devereux and Salvanes (2007) find a 10% increase in birth weight to increase the probability of high school completion by 1 percentage point and to increase birth weight for the next generation by 1.5%. The 1.5% lower birth weight effect of capitation contracts therefore translates into 0.15 percentage point lower probability of high school completion and a 0.23% decrease in the birth weight for the next generation. Using a sample of California births 1960-1982, Royer (2009) finds the effect of birth weight on high school completion to be half the size of the Norwegian study. This means that the effect of capitation system would translate into 0.07 percentage point decrease in high school completion, if similar effects of a change in remuneration can be found in the U.S.

Physician-induced Demand.—Many studies suggest the existence of physician-induced demand. Investigating the changes in the number of GP activities in Copenhagen compared to Copenhagen county, Krasnik et al. (1990) find that the average increase in diagnostic and curative services is 61%. Simultaneously, they find referrals to specialists and hospitals decreasing by 9.5%. To estimate the net inducement in Copenhagen compared to

Copenhagen County, I adjust the number of activities according to the number of specialist referrals as presented by Krasnik et al. (1990). The average number of services provided per patient per week is 1.5 [8], and the average number of specialist visits is 0.7 per patient. Assuming that one referral directly translates into 1.5 fewer activities per consultation and 0.7 fewer GP visits per patient (The National Board of Health 2003), the average decrease in referrals to specialists and hospitals found in Krasnik et al. (1990) should be multiplied by 2.5, i.e. 23.8% in total. The average increase in activities is therefore 61% whereas the average decrease in referrals is 24% leading to a rough estimate of the net inducement of 37%.

This inducement rate is somewhat similar to what is found in other studies. Gruber and Owings (1996) estimate the effect of inducement to be 16-32%. Although, using Canadian data Devlin & Sarma (2008) estimate the increase in GP visits to 58% when taking self-selection into the remuneration system into account. However, if I assume that the total level of inducement translates directly into inducement in prenatal care, the 37% net inducement rate will translate into a 1.5% increase in birth weight.

Comparing the Effects of Remuneration to Effects of Parental Schooling.—A different branch of the literature investigates infant health as a social return to maternal schooling. While the effect of maternal schooling on infant health to some extent is inconclusive, causal effects of maternal schooling on birth weight are found in the U.S. by Currie & Moretti (2003), in the U.K. by Chevalier and O'Sullivan (2007), and in Denmark by Bingley, Christensen & Jensen (2009).

On a U.S. sample of births, Currie and Moretti (2003) find 0.9 years of schooling to increase the probability of low birth weight by 2% and the probability of preterm birth by 1% using college openings to generate exogenous variation in maternal schooling. Similar local average treatment strategy is applied by Chevalier and O'Sullivan (2007) who estimate the effect of maternal schooling on birth weight in the U.K. Chevalier and O'Sullivan

[8] This number is estimated using data on health services from 1997 to 1999, using a 5% random sample of the population. The estimates condition on the patient having at least one GP visit, i.e. all patients not having any GP visits are deleted from the sample.

find that the effect of one additional year of schooling increases birth weight by 2-6%. Although, these effects are estimated in the lower part of the maternal education distribution (leaving school at age 14-15), and using older cohorts (1948). While the U.S. and the U.K. examples target different ends of the maternal education distribution, Bingley, Christensen & Jensen (2009) estimate the average treatment effect of maternal schooling using an intergenerational within-twin-pair estimation strategy. They find one additional year of maternal schooling to increase the average birth weight by 0.9% using Danish infants born 1978-2001.

Combining the previous findings of maternal schooling on birth weight to that of the remuneration changes, the results suggest that the effect of remuneration is around half the size of one additional year of schooling in the British study, two thirds of one year of additional schooling in the U.S. study and one and three quarters of the effect of one additional year of maternal schooling in Denmark.

In total, these approximate calculations, which should be interpreted with caution, suggest that the cost containment of capitation contracts compared to the mixed system (where inducement was increased by 37%) needs to be adjusted for the long-run effects of infant.

10. REFERENCES

Almond, D., Chay, K. Y., & Lee, D. S. 2005, "The Costs of Low Birth Weight", Quarterly Journal of Economics, vol. 120, no. 3, pp. 1031-1083.

Anand, D., Stevenson, C. J., West, C. R., and Pharoah, P. O. D. 2008, "Lung function and respiratory health in adolescents of very low birth weight", Achieves of Disease in Childhood, vol 91, no 1, pp. 101-106.

Barker, D. J. P., Osmond, C. Rodin, I., Fall, C. H. D., Winter, P. D. 1995, "Low weight gain in infancy and suicide in adult life", British Medical Journal, vol 311, no 1203.

Becker, G. S. 1965, "A Theory of the Allocation of Time", The Economic Journal, vol. 75, no. 299, pp. 493-517.

Behrman, J. R. & Rosenzweig, M. R. 2004, "Returns to Birthweight", Review of Economics & Statistics, vol. 86, no. 2, pp. 586-601.

Bertrand, M., Duflo, E., & Mullainathan, S. 2004, "How Much Should We Trust Differences-in-Differences Estimates?", Quarterly Journal of Economics, vol. 119, no. 1, pp. 249-275.

Bingley, P., Christensen, K., & Jensen, V. M. 2009, Parental Schooling and Child Development: Learning from Twin Parents. SFI working paper series. WP no 07:2009

Black, S. E., Devereux, P. J., & Salvanes, K. G. 2007, "From the Cradle to the Labor Market - The effect of Birthweight on Adult Outcomes", Quarterly Journal of Economics, vol. 122, no. 1, pp. 409-439.

Blundell, R. & Costa Dias, M. 2000, "Evaluation Methods for Non-Experimental Data", Fiscal Studies, vol. 21, no. 4, pp. 427-468.

Chevalier, A. & O'Sullivan, V. 2007, "Mother's Education and Birth Weight", IZA Discussion Paper Series no. 2640.

Currie, J. 2009, "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development", Journal of Economic Literature, forthcoming

Currie, J. & Gruber, J. 1996, "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women", Journal of Political Economy, vol. 104, no. 6, pp. 1263-1296.

Currie, J. & Moretti, E. 2003, "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings", *Quarterly Journal of Economics*, vol. 118, no. 4, pp. 1495-1532.

Del Bono, E. Ermisch, J. & Francesconi, M. 2008, "Intrafamily Resource Allocations: A Dynamic Model of Birth Weight", IZA Discussion Paper Series no. 3704.

Devlin, R. A. & Sarma, S. 2008, "Do physician remuneration schemes matter? The case of Canadian family physicians", Journal of Health Economics, vol. 27, no. 5, pp. 1168-1181.

Dranove, D. & Wehner, P. 1994, "Physician-induced demand for childbirths", Journal of Health Economics, vol. 13, no. 1, pp. 61-73.

Evans, R. G. 1974, "Supplier-induced demand: Some empirical evidence and implications," in The economics of health and medical care, M. Perlman, ed., Macmillan, London & Basingstoke, pp. 162-173.

Fuchs, V. R. 1978, "The Supply of Surgeons and the Demand for Operations", American Economic Review, vol. 86, pp. 35-56.

Gosden, T., Forland, F., Kristiansen, I., Sutton, M., Leese, B., Guiffrida, A., Sergison, M., & Pedersen, L. Capitation, salary, fee-for-service and mixed systems of payments: effects on the behaviour of primary care physicians. Cochrane Database of Systematic Reviews [3]. 2000.

Gosden, T., Pedersen, L., & Torgerson, D. 1999, "How should we pay doctors? A systematic review of salary payments and their effect on doctor behaviour", QJM, vol. 92, no. 1, pp. 47-55.

Gruber, J. & Owings, M. 1996, "Physician Financial Incentives and Cesarean Section Delivery", The RAND Journal of Economics, vol. 27, no. 1, pp. 99-123.

Joergensen, F. S. 2002: "Obstetric Ultrasound Screening of Pregnant Women in Denmark 2000 - with a description of the development since 1990", [in Danish] Journal of the Danish Medical Association, vol. 165, no. 46, pp 4404-4409.

Joergensen, C. 2003:"The Possibility of Discovering Congenital Defects before Birth by Means of Prenatal Ultrasound Screening", [in Danish] Journal of the Danish Medical Association, vol. 164, no. 48, pp 5603-5607.

Kouides, R. W., Bennett, N. M., Lewis, B., Cappuccio, J. D., Barker, W. H., LaForce, F. M., & the primary-care physicians of Monroe County 1998, "Performance-based physician reimbursement and influenza immunization rates in the elderly", American Journal of Preventive Medicine, vol. 14, no. 2, pp. 89-95.

Krasnik, A., Groenewegen, P. P., Pedersen, P. A., von Scholten, P., Mooney, G., Gottschau, A., Flierman, H. A., & Damsgaard, M. T. 1990, "Changing remuneration systems: effects on activity in general practice", BMJ: British Medical Journal, vol. 300, no. 6741, pp. 1698-1701.

Krasnik, A., Scholten, P. V., & Damsgaard, M. T. 1991, "General Practitioners' Attitudes to a Recent Change in their Remuneration System", Scandinavian Journal of Primary Health Care, vol. 9, no. 2, pp. 83-87.

McGuire, T. G. 2000, " Physician Agency," in Handbook of Health Economics, Volume 1, Part 2 edn, J. C. a. J. Anthony, ed., Elsevier, pp. 1175-1200.

McGuire, T. G. & Pauly, M. V. 1991, "Physician response to fee changes with multiple payers", Journal of Health Economics, vol. 10, no. 4, pp. 385-410.

Ministry of the Interior and Health 1985, Guidelines for Midwifery and Care during Pregnancy [in Danish] Ministry of Health, Copenhagen.

Ministry of the Interior and Health. Executive Order # 748 of 16/08. Regarding EU citizens access to practicing Medical Care, Dentistry and Midwifery. 748 of 1608 1994. 1994. Ref Type: Bill/Resolution

Mullen, K., Frank, R.B., Rosenthal, M. (2009): "Can You Get What You Pay For? Pay-for-Performance and the Quality of Health Care Providers", National Bureau of Economic Research Working Paper Series, working paper no. 14886.

National Health Security System & Association of General Practitioners. Agreement on General Practice. 1984. Ref Type: Bill/Resolution

Ohlsson, A. & Shah, P. 2008, *Determinants and Prevention of Low Birth Weight: A Synopsis of the Evidence* The Institute of Health Economics (IHE) , Alberta, Canada.

Oreopoulos, P., Page, M. E., & Stevens, A. H. 2006, "The Intergenerational Effects of Compulsory Schooling", Journal of Labor Economics, vol. 24, no. 4, pp. 729-760.

Phelps, C. E. 1986, "Induced demand – can we ever know its extent?", Journal of Health Economics, vol. 5, no. 4, pp. 355-365.

Power, C., Bartley, M., Smith GD., Blane D 1996 "Transmission of Social and biological risk across the life course", in Health and Social Organisation, ed D. Blane, E. Brunner, & R. Wilkinson. pp.188-203. New York: Routledge

Royer, H. 2009, "Separated at Girth: US Twin Estimates of the Effects of Birth Weight", American Economic Journal: Applied Economics, vol. 1, no. 1, pp. 49-85.

Scott, A. 2000, "Economics of general practice," in Handbook of Health Economics, Volume 1, Part 2 edn, J. C. a. J. Anthony, ed., Elsevier, pp. 1175-1200.

Scott, A. & Hall, J. 1995, "Evaluating the effects of GP remuneration: problems and prospects", Health Policy, vol. 31, no. 3, pp. 183-195.

Sugiura-Ogasawarna M., Furukawa T., Nakano, Y. Hori, S. Aoki, K., & Kitamura T. 2002, "Depression as a Potential Causal Factor in Subsequent Miscarriage in Current Spontaneous Aborters", Human Repordution, vol. 17, no. 10 pp. 2580-2584.

The National Board of Health 2003, Contacts to General Practitioners under National Health Service [in Danish], National Board of Health, Copenhagen.

The National Board of Health 2007a, "Birth at Hospitals and the Size of the Maternity Wards [in Danish]", New Figures from the National Board of Health, vol. 11, no. 3.

The National Board of Health 2007b, "The Use of Anti-Depressant Pills in Relation to Pregnancy and Birth 1997-2006 [in Danish]", New Figures from the National Board of Health, vol. 11, no. 20.

Town, R., Kane, R., Johnson, P., & Butler, M. 2005, "Economic incentives and physicians' delivery of preventive care: A systematic review", American Journal of Preventive Medicine, vol. 28, no. 2, pp. 234-240.

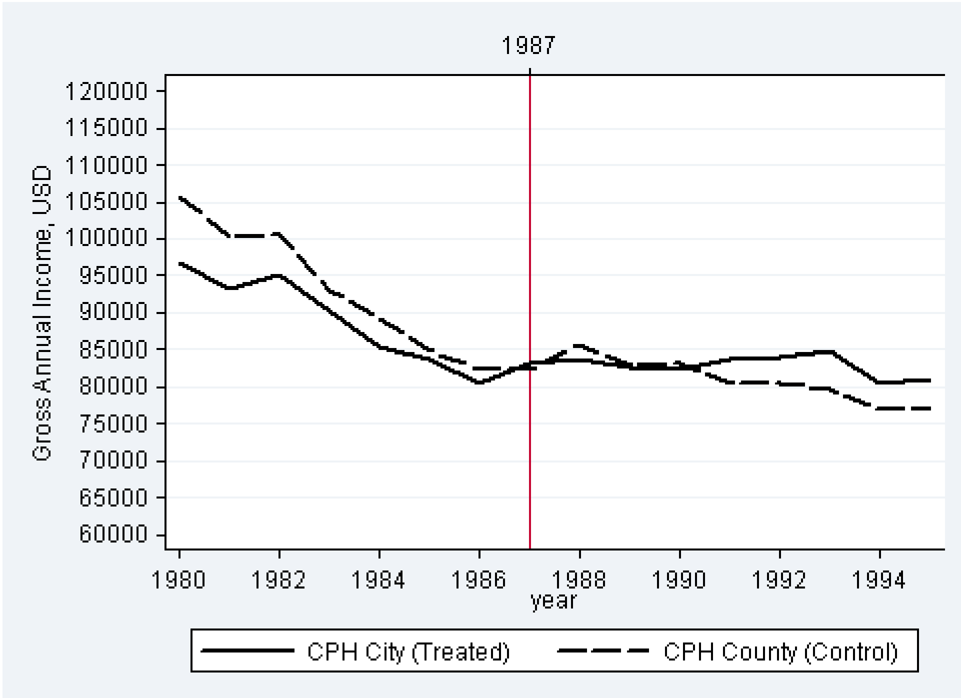
Vallgaarda, S., Krasnik, A., Vrangbaek, K. 2001, Health Care Systems in Transition, Denmark. WHO : European Observatory on Health Care Systems, vol 3 no 7.

Vedsted, P., Olesen, F., Hollnagel, H., Bro, F., & Kamper-Jørgensen, F. 2005, General Practice in Denmark- [in Danish] Journal for General Practice & The Research Unit for General Practice, Aarhus, Copenhagen.

Woodward, R. S. & Warren-Boulton, F. 1984, "Considering the effects of financial incentives and professional ethics on the 'appropriate' medical care", Journal of Health Economics, vol. 3, no. 3, pp. 223-237.

World Health Organisation & The United Nations Children's Fund 2004, Low Birthweight, WHO Publications.

Figure 1: GP's Gross Annual Income in Treatment and Control Areas 1980



Source: Statistics Denmark, Integrated Database for Labour Market Research.

Tabulation of gross annual income is conditional on GP age and tenure, and county-, municipality-, and church-tax. GP income is found using the workplace identifier from 1992, because GP workplace identifiers are pooled with all other types of health practice until 1992. I track back GP workplace identifiers from 1992 to 1980, and use income from all persons with a medical degree working at the GP workplace in the given year. In total I have 22650 person years. Estimating the GP income difference in Copenhagen city 5 years before and after 1987 relative to Copenhagen county, income changes 7%, significant at a 99% level.

Table 1: Summary Statistics for Treated and Controls Groups

	Copenhagen City (treated)			Copenhagen County (controls)		
	Before	After	Diff.	Before	After	Diff.
Birth weight	3308.7	3365.4	56.7	3332.6	3334.9	2.3
	(557.6)	(575.4)	(-17.8)	(550.5)	(590.1)	(-39.6)
Low birth weight	0.060	0.057	0.00	0.062	0.063	0.00
N	[2382]	[3113]		[1765]	[1979]	
Preterm birth	0.06	0.05	-0.01	0.043	0.061	0.02
N	[2378]	[3113]		[1760]	[1980]	
Foetal growth	83.45	84.77	1.32	84.21	84.30	0.08
	(12.35)	(12.92)	(-0.57)	(12.29)	(13.16)	(-0.87)
N	[2378]	[3099]		[1760]	[1970]	
One min. APGAR score	0.52	0.54	0.02	0.6	0.524	-0.08
Five min. APGAR score	0.92	0.91	-0.01	0.93	0.90	-0.03
N	[2383]	[3130]	[747]	[1765]	[1994]	

Means and standard deviations are reported. Standard deviations are in parenthesis and the number of observations are in brackets. Low birth weight is defined as below 2500 gram, preterm birth as born before 37th week, foetal growth as grams per week. APGAR scores are coded as a dummy for score equal 10 (optimal health condition).

Table 1: Summary Statistics for Treated and Controls Groups *continued*

	Copenhagen City (treated)			Copenhagen County (controls)		
	Before	After	Diff.	Before	After	Diff.
<u>Covariates (individual level)</u>						
Mother's age	26.30	26.97	0.68	25.66	26.57	0.92
	(4.36)	(4.22)	(0.14)	(4.394)	(4.451)	(-0.06)
Ethnic Danes	0.87	0.83	-0.04	0.894	0.837	-0.06
Mother's schooling	11.93	12.36	0.42	11.89	12.25	0.36
	(2.66)	(2.47)	(0.18)	(2.533)	(2.497)	(0.04)
Father's schooling	12.19	12.55	0.36	12.32	12.56	0.24
	(2.95)	(2.80)	(0.15)	(2.90)	(2.75)	(0.15)
Mother's log income	3.96	4.13	0.16	3.95	4.27	0.33
(in 1000 DKK)	(2.59)	(2.79)	(-0.2)	(2.76)	(2.78)	(-0.02)
Father's log income	4.45	4.54	0.10	4.80	4.89	0.07
(in 1000 DKK)	(2.09)	(2.48)	(-0.39)	(1.68)	(2.17)	(-0.49)
Married	0.34	0.35	0.01	0.41	0.44	0.03
Single	0.22	0.20	-0.01	0.17	0.15	-0.02
Cohabitant	0.44	0.45	0.00	0.43	0.42	-0.01
Male infant	0.52	0.51	-0.02	0.52	0.53	0.02
<u>GP (municipality averages)</u>						
GP age squared	6.68	6.85	0.16	6.61	6.82	0.21
	(0.08)	(0.14)	(-0.06)	(0.35)	(0.31)	(0.04)
GP age	44.69	46.92	2.23	43.86	46.64	2.78
	(1.12)	(1.93)	(-0.81)	(4.66)	(4.22)	(0.44)
GP gender	0.57	0.54	-0.04	0.56	0.59	0.03
	(0.11)	(0.10)	(0.00)	(0.15)	(0.15)	(-0.01)
N	[2210]	[2928]		[1687]	[1904]	

Standard deviations are reported in parenthesis and the number of observations is listed in brackets. Parental income is reported as average log income, but converted into yearly quartiles in the regression.

Table 2: Services Provided during Prenatal Care

	Women, pregnant		Price per Item	Women, not pregnant	
	Mean	Std dev.	USD	Mean	Std dev.
Face to face consultation	6.16	3.36	5.67	1.80	2.64
Telephone consultation	3.45	2.72	3.12	1.19	2.13
Blood sample	0.59	0.63	4.87	0.00	0.000
Urine sample	1.44	0.76	4.87	0.00	0.04
Urine lab test	2.41	2.08	1.23	0.17	0.77
Blood lab test	1.52	1.20	1.23	0.07	0.37
Culturing bacteria	1.74	1.24	3.71	0.06	0.36
N		[3042]			[80286]
Total average fee (USD)		66.20			14.29
Yearly capitation fee per patient after 1987		26.88			

Mean and standard deviation of number of services is estimated using a 5 % random sample of women born 1964-1977 during 1997-1999. The pregnant women are all carrying their firstborn child and number of services is measured in a 10-month period before giving birth. For the group of women not being pregnant, the number of services is measured in 10-months period per year. Urine and blood lab tests compound different types of tests. Given prices varies per test, the lowest price per item is used. Blood analyses include hepatitis, hemoglobin testing, and gestational diabetes; urine tests include pregnancy test and pre-eclampsia test; and culturing bacteria primarily includes cystitis testing. The listed prices per item stems from the National Health Security System and the Organization of General Practitioners (1984).

Figure 2: Outcome Distribution for Treated and Controls, by Maternal Age and Policy Period

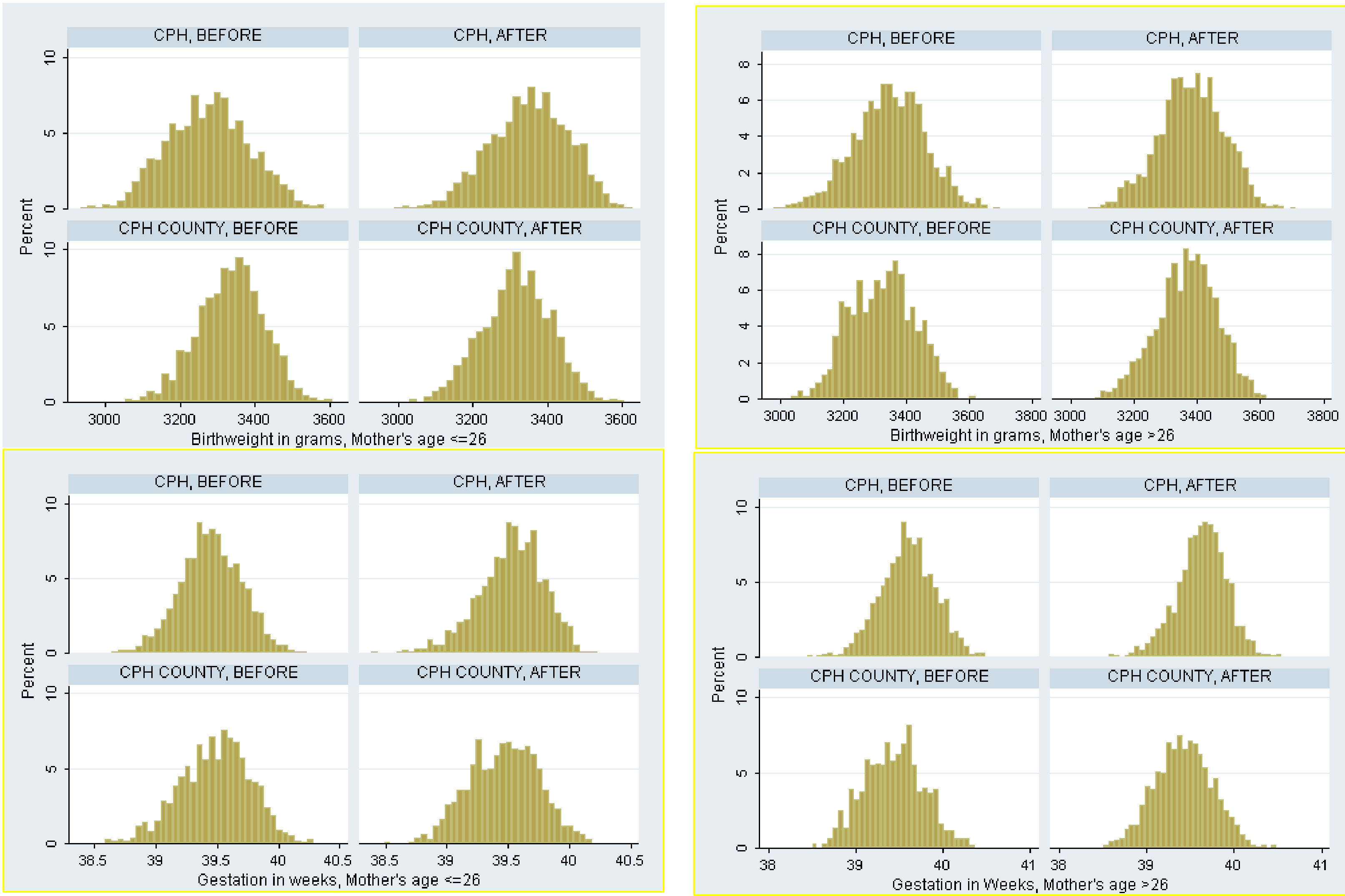


Figure 2: Outcome Distribution for Treated and Controls, by Maternal Age and Policy Period *continued*

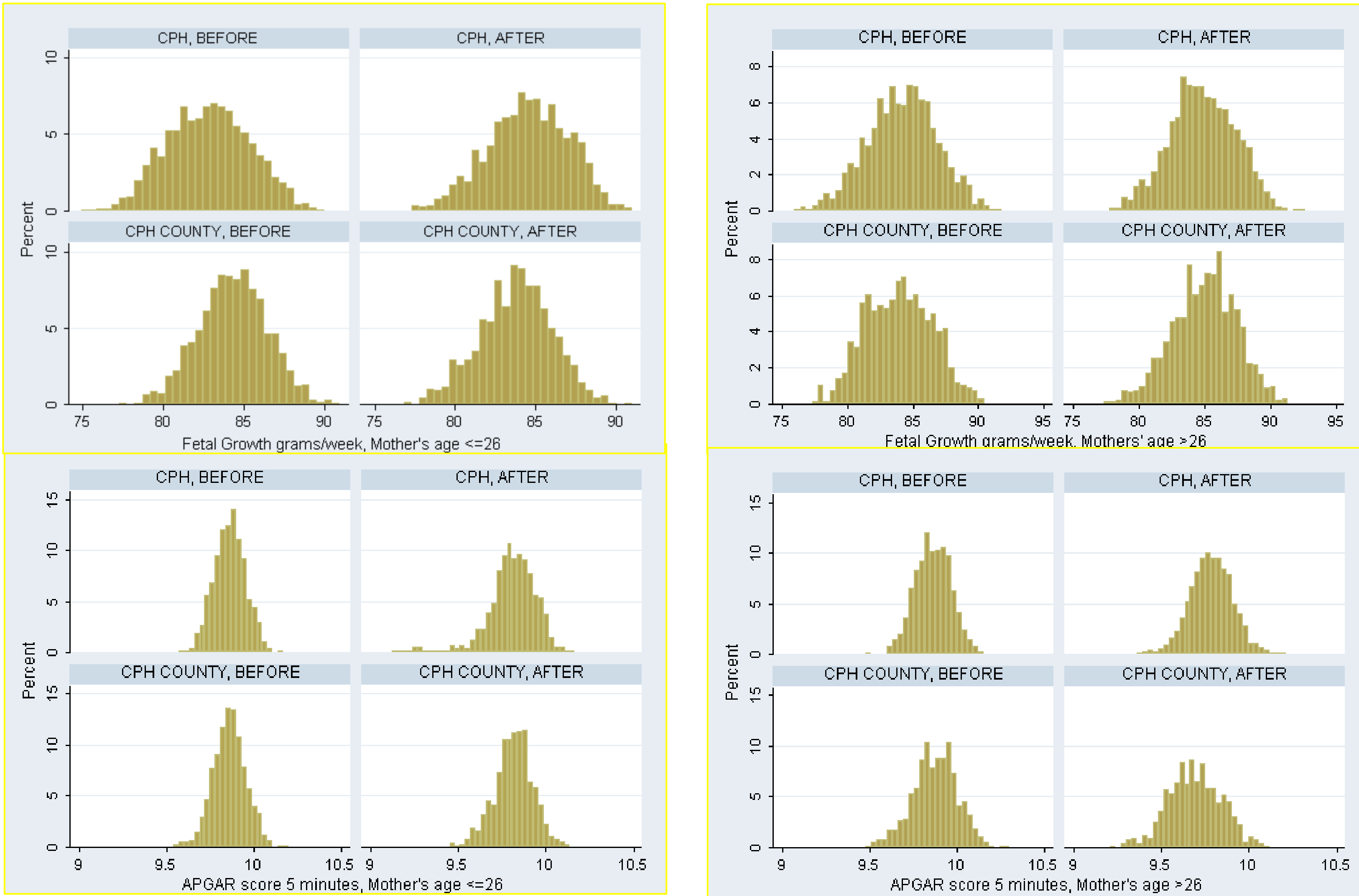
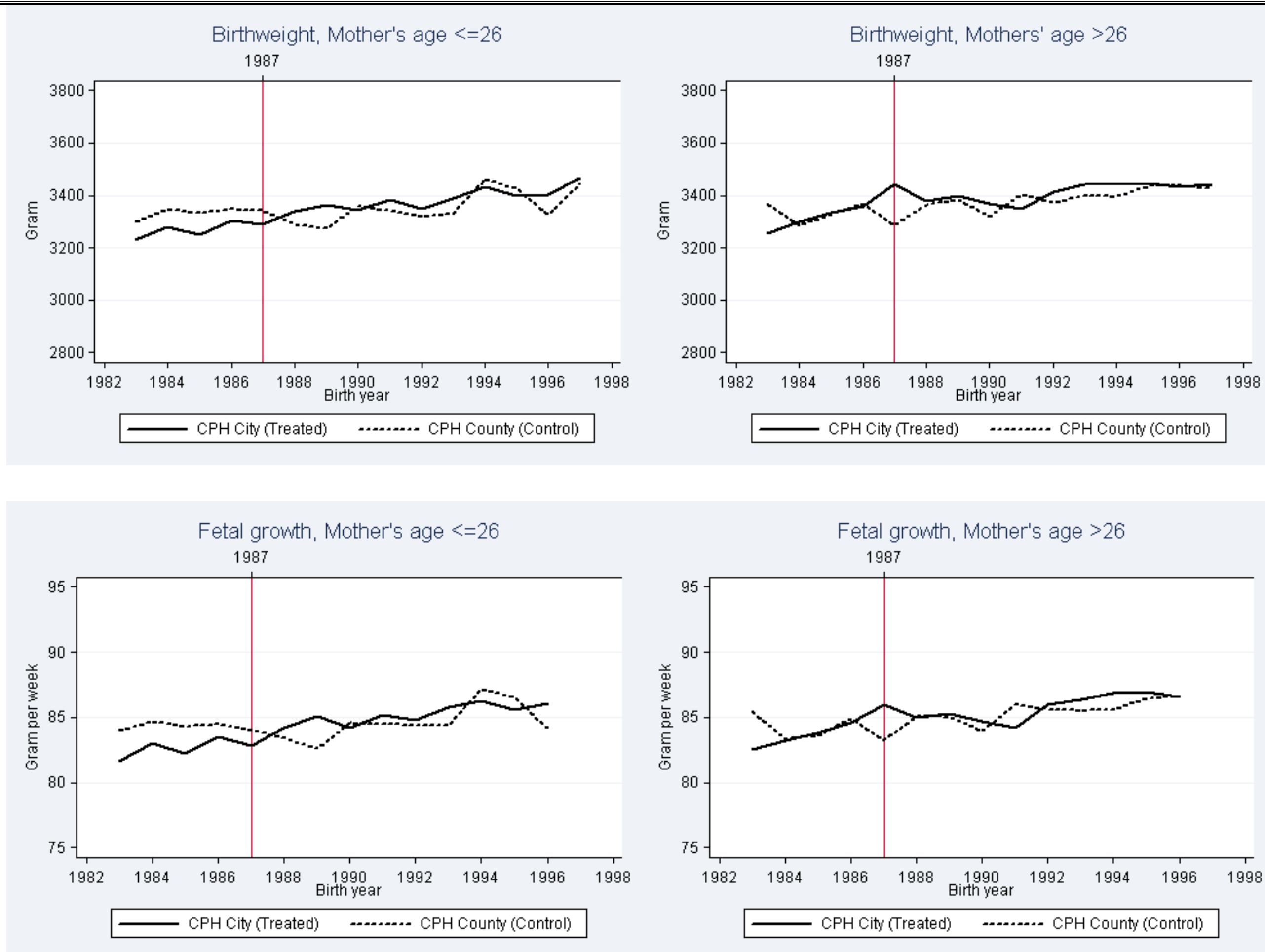
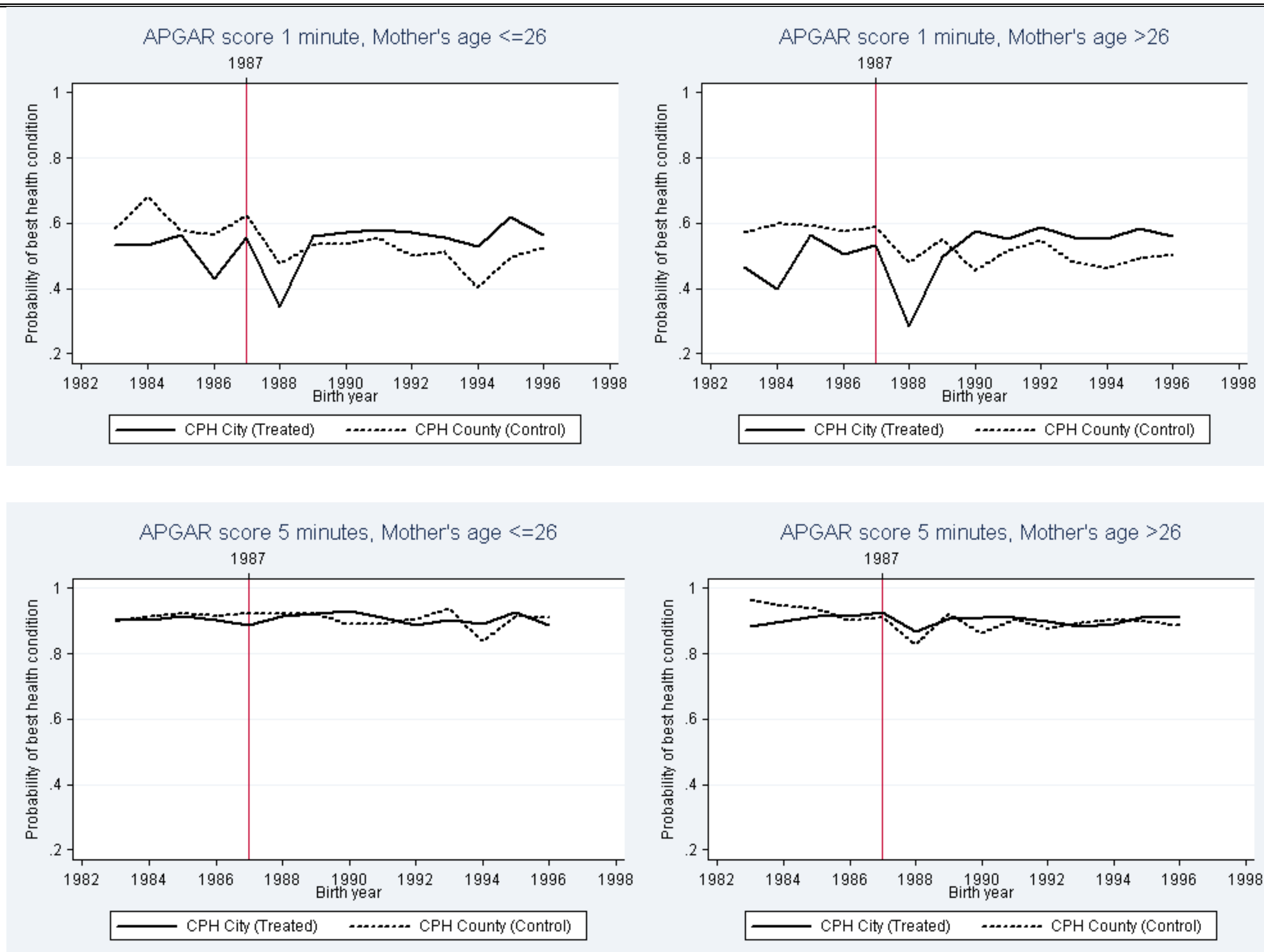


Figure 3: Outcome Distribution across Birth Years for Treated and Controls, by Maternal Age



The graphs are computed as municipality fixed effect regressions including all covariates as listed in table 1.

Figure 3: Outcome Distribution across Birth Years for Treated and Controls, by Maternal Age *continued*



The graphs are computed as municipality fixed effect regressions including all covariates as listed in table 1.

Table 3: The Effect of Capitation Contracts on Infant Health

	All Mothers		Mothers, age<=26		Mothers, age>26	
	No Covariates	With Covariates	No Covariates	With Covariates	No Covariates	With Covariates
Birth weight	-53.7 **	-50.8 **	-89.3 ***	-80.6 **	1.1	-4.2
	(24.3)	(24.5)	(32.2)	(32.6)	(37.3)	(38.0)
Low birth weight	0.005	0.003	0.014	0.013	-0.010	-0.011
	(0.010)	(0.010)	(0.014)	(0.014)	(0.015)	(0.016)
N	[9239]		[4932]		[4307]	
Preterm birth	0.023 **	0.021 **	0.026 **	0.027 **	0.014	0.011
	(0.010)	(0.010)	(0.013)	(0.013)	(0.015)	(0.016)
N	[9231]		[4925]		[4306]	
Foetal growth	-1.229 **	-1.158 **	-2.163 ***	-1.953 ***	0.098	0.013
	(0.543)	(0.547)	(0.723)	(0.729)	(0.831)	(0.845)
N	[9207]		[4916]		[4219]	
One minute APGAR score	0.000 ***	-0.094 ***	-0.095 ***	-0.097 ***	-0.089 ***	-0.094 ***
	(0.021)	(0.022)	(0.029)	(0.029)	(0.032)	(0.033)
Five minutes APGAR score	-0.020	-0.022 *	-0.013	-0.014	-0.030	-0.034 *
	(0.012)	(0.012)	(0.016)	(0.016)	(0.018)	(0.019)
N	[9272]		[4943]		[4329]	

* p<0.10,**p<0.05,***p<0.01. Standard errors are listed in parenthesis and the number of observations is listed in brackets. Low birth weight is defined as below 2500 gram, preterm birth as born before 37th week, foetal growth as grams per week. APGAR scores are coded as a dummy for score=10 (optimal health condition). Each estimate represents a separate regression and includes municipality fixed effects.

Table 4: The Effect of Capitation on Infant Health, 3 Year Window

	All Mothers	Mothers age<=26	Mothers age>26
Birth weight	-33.9	-81.3 *	30.1
	(33.4)	(44.2)	(51.9)
Low birth weight	-0.006	0.002	-0.015
	(0.014)	(0.019)	(0.021)
N	[4934]	[2704]	[2230]
Preterm birth	0.016	0.022	0.006
	(0.014)	(0.018)	(0.021)
Foetal growth	-0.896	-2.185 **	0.767
	(0.745)	(0.990)	(1.152)
N	[4927]	[2698]	[2229]
One minute APGAR score	-0.055 *	-0.069 *	-0.034
	(0.029)	(0.040)	(0.045)
Five minutes APGAR score	-0.031 *	-0.041 *	-0.026
	(0.017)	(0.022)	(0.026)
N	[4934]	[2704]	[2230]

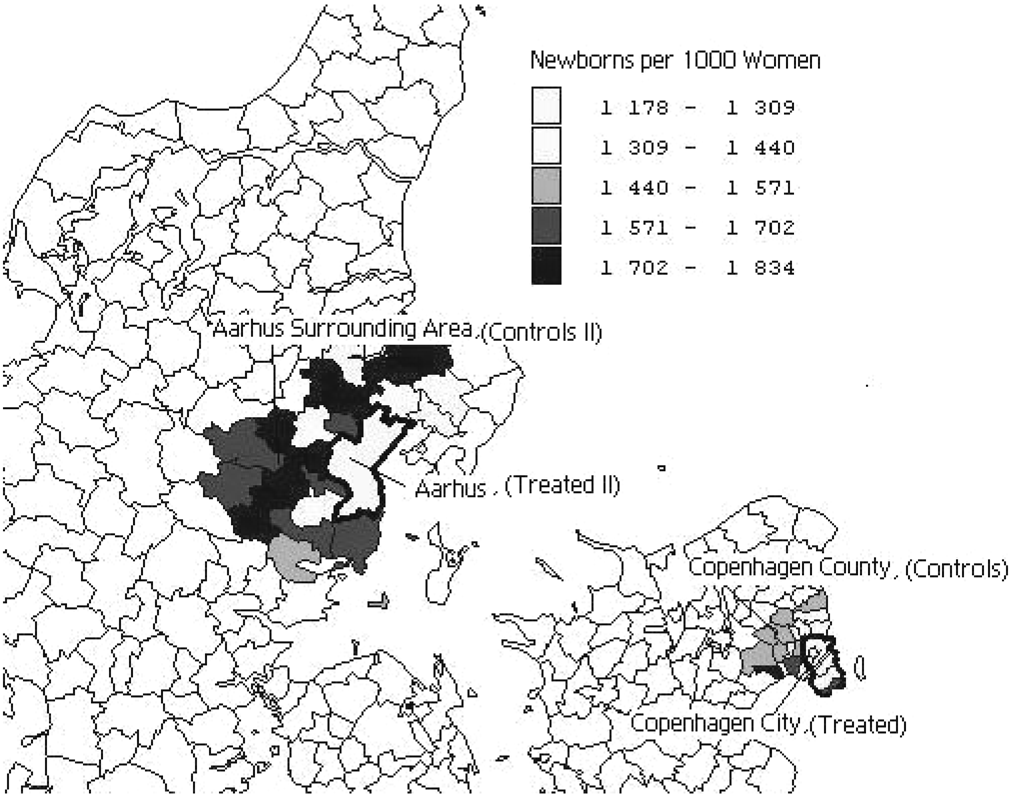
* p<0.10,**p<0.05,***p<0.01. Standard errors are listed in parenthesis and the number of observations is listed in brackets. Low birth weight is defined as below 2500 gram, preterm birth as born before 37th week, foetal growth as grams per week. APGAR scores are coded as a dummy for score=10 (optimal health condition). Each estimate represents a separate regression and includes municipality fixed effects.

Table 5: The Effect of Capitation on Infant Health, Time Trend Adjusted

	All Mothers	Mothers, age<=26	Mothers, age>26
Birth weight	-77.2 *	-119.7 *	-17.5
	(43.7)	(61.2)	(64.0)
Low birth weight	0.004	0.02	-0.008
	(0.017)	(0.025)	(0.025)
N	[12153]	[5618]	[6535]
Preterm birth	0.039 **	0.021	0.05 *
	(0.018)	(0.025)	(0.026)
N	[12128]	[5606]	[6522]
Foetal growth	-1.403	-2.977 **	0.492
	(0.972)	(1.372)	(1.419)
N	[12068]	[5583]	[6485]
One minute APGAR score	-0.042	-0.039	-0.042
	(0.038)	(0.055)	(0.055)
Five minutes APGAR score	0.003	0.000	0.012
	(0.022)	(0.031)	(0.033)
N	[12228]	[5643]	[6585]

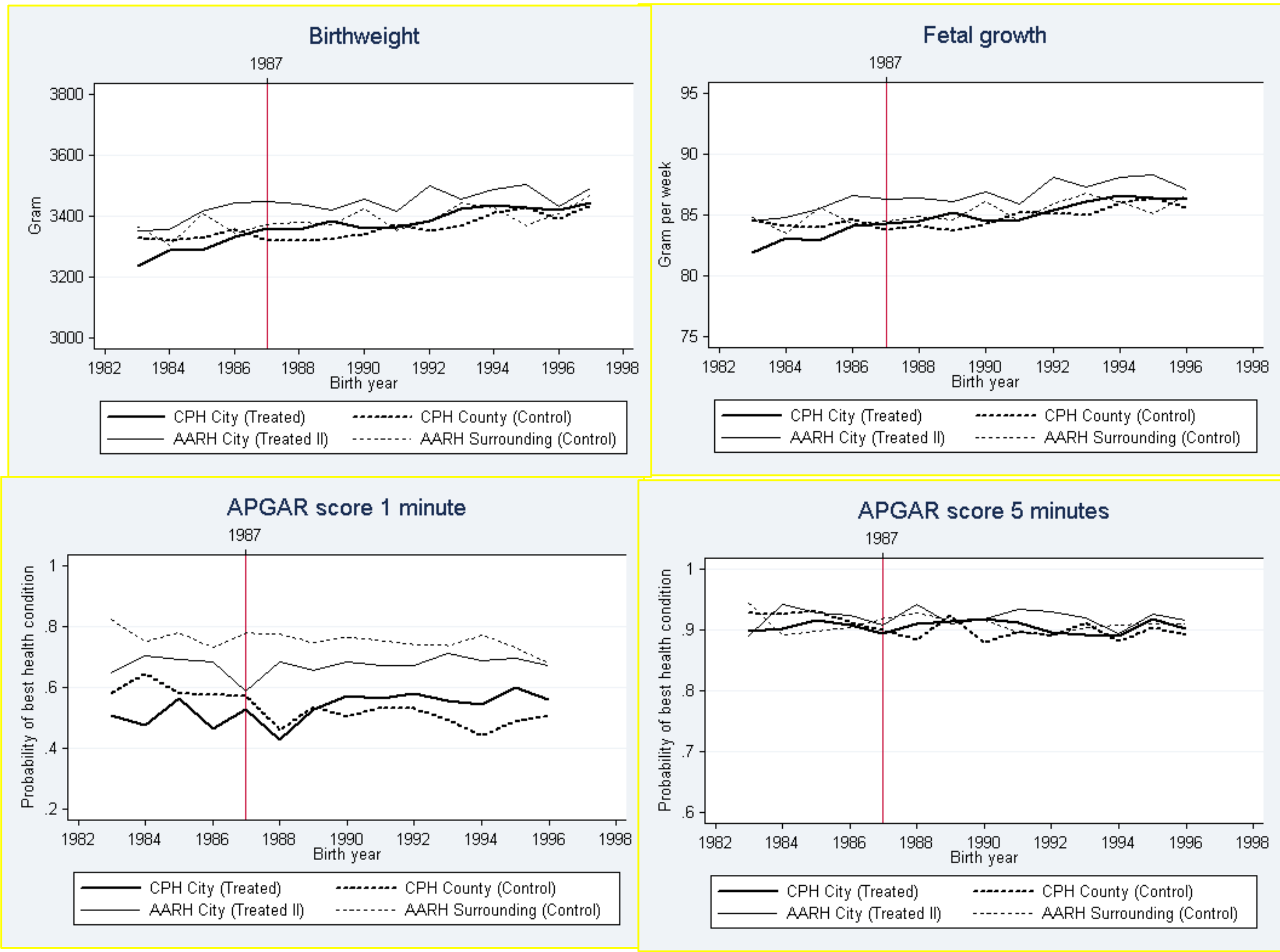
* p<0.10,**p<0.05,***p<0.01. Standard errors are listed in parenthesis and the number of observations is listed in brackets. Low birth weight is defined as below 2500 gram, preterm birth as born before 37th week, foetal growth as grams per week. APGAR scores are coded as a dummy for score=10 (optimal health condition). Each estimate represents a separate regression and includes municipality fixed effects.

Figure 4: Fertility Ratio in Treated and Control groups, Birth Year 1987



Source: Statistics Denmark, Dynamic Databases 2009

Figure 5: Infant Health Outcomes Distributed across Birth Year for Treatment and Control groups



The figures are computed using municipality fixed effect regressions, including all covariates as listed in Table 1.

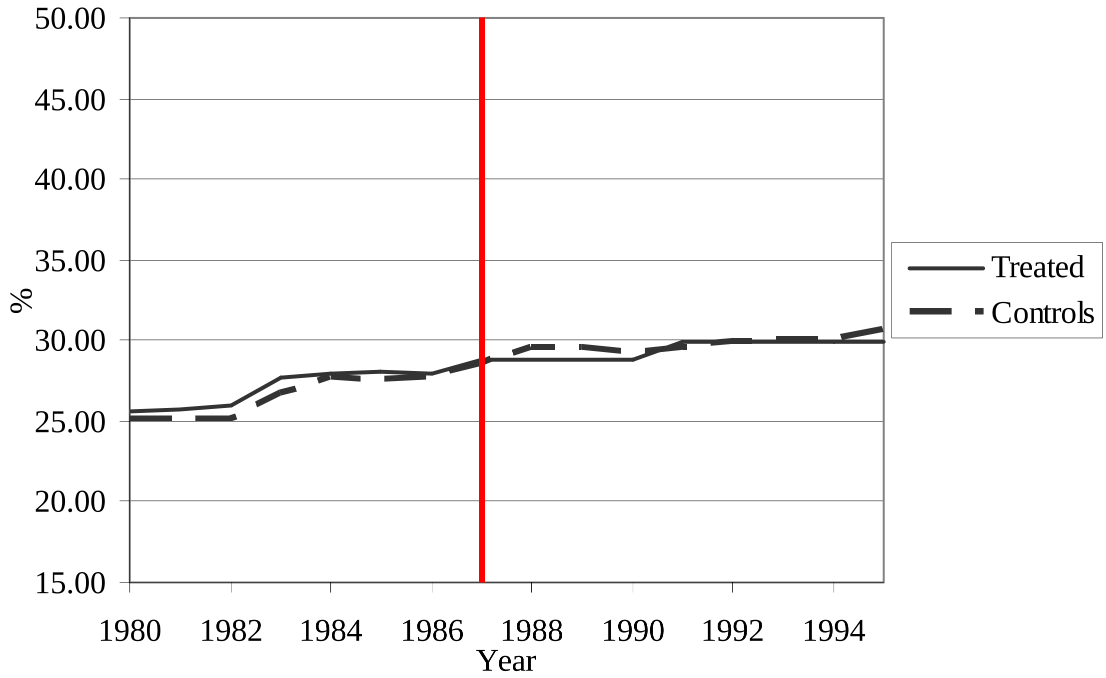
Table 6: The Effect of Capitation on Infant Health, adjusted for Inner City Residence

	All Mothers	Mothers, age<=26	Mothers, age>26
Birth weight	-45.2	-100.7 *	8.4
	(41.8)	(55.1)	(67.0)
Low birth weight	-0.017	0.008	-0.050 *
	(0.017)	(0.023)	(0.027)
N	[13708]	[7407]	[6301]
Preterm birth	0.012	0.029	-0.012
	(0.017)	(0.022)	(0.027)
N	[13686]	[7393]	[6293]
Fetal growth	-1.059	-2.420 *	0.215
	(0.934)	(1.237)	(1.493)
N	[13654]	[7380]	[6274]
One minute APGAR score	0.072 **	0.088	0.071
	(0.036)	(0.049)	(0.058)
Five minutes APGAR score	0.023	0.006	0.046
	(0.021)	(0.028)	(0.034)
N	[13750]	[7422]	[6328]

* p<0.10,**p<0.05,***p<0.01. Standard errors are listed in parenthesis and the number of observations is listed in brackets. Low birth weight is defined as below 2500 gram, preterm birth as born before 37th week, foetal growth as grams per week. APGAR scores are coded as a dummy for score=10 (optimal health condition). Each estimate represents a separate regression and includes municipality fixed effects.

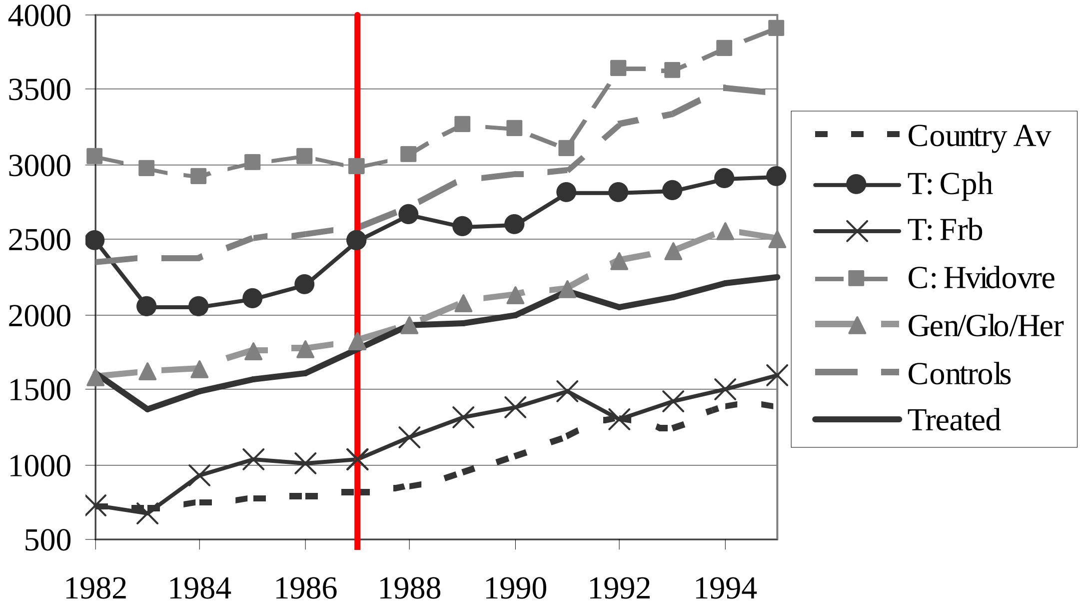
APPENDIX

Figure A1: Regional Taxation in Copenhagen city and Copenhagen County



Source: Statistics Denmark, Dynamic Databases 2009. The tax levels are the sum of county, municipality, and church tax.

Figure A2: Number of Hospital Births in The Treatment and Control Areas



Source: The National Board of Health (2007A). I compute an average of the number of births at hospitals in Gentofte, Glostrup, and Herlev (all control areas), because the number of hospital births is similar at these hospitals.