Essays on the Economics of Health and Fertility

Karin Monstad

June 2007

Dissertation submitted to the Department of Economics, the Norwegian School of Economics and Business Administration, in partial fulfilment of the requirement for the degree Ph.D.

Acknowledgements

First and foremost I am indebted to my advisory committee: Professor Kjell G. Salvanes, Professor Fred Schroyen, Professor Carol Propper and Professor Jan Erik Askildsen. With his patience and devotion to the task, my main advisor at the outset, Fred Schroyen, has been a great support throughout the work with this dissertation. As my field of interest changed into purely empirical work, Kjell G. Salvanes took over as main advisor. Kjell's energy, encouragement and clear advice have been most inspiring. Together with Carol Propper, Kjell co-authored the paper "Education and Fertility: Evidence from a Natural Experiment". Their precise comments and suggestions have improved my research immensurably.

The empirical work has benefited from data from various sources. The paper "Patients' Preferences for Choice of Hospital" uses data that originate from The Norwegian Arthroplasty Register. I am grateful to Jan Erik for introducing me to this interesting data register and for helping me meet the challenges involved in research cooperation across different disciplines. Working with Birgitte Espehaug and Lars Birger Engesæter has been a valuable learning experience.

I gratefully acknowledge the Ph.D. scholarship from the Norwegian School of Economics and Business Administration and financial support from the Health Economics Bergen Programme. Together, these institutions have provided a stimulating academic and social environment.

Finally, my thanks go to all friends and relatives who endured my frustrations while working on this dissertation. Special thanks to my sister and to Inge and our sons, Dag Inge and Håkon.

Bergen, June 2007.

Karin Monstad

Table of contents

Introduction

What can economics bring to the study of such diverse phenomena as the choice of hospital for elective operations and individuals' decisions on fertility? In this dissertation, both subjects are treated using an approach that is common to most studies within the field: analysing trade-offs, assuming rational choices and utility-maximizing behaviour. Human capital is a central concept in all three analyses in this dissertation. I will first give a brief presentation of this analytical tool, and second, explain how it relates to my work.

In a narrow sense, human capital refers to the productive capacities of human beings as income-producing agents in the economy. Education is the most common example of investment in human capital. However, I will use human capital in a broader sense that also includes a person's health endowment. Human Capital Theory emerged in the 1960s and 1970s and the fundamental conceptual framework was provided by Gary Becker (1964). Becker describes it as follows: "Human capital analysis starts with the assumption that individuals decide on their education, training, medical care, and other additions to knowledge and health by weighing the benefits and costs" (Becker, 1993, p. 392). The theory has applied well-known concepts like investment, rate of return and depreciation in a novel way and has provided explanations of human behaviour in a number of fields, including fertility and the demand for health services.

In his Nobel Lecture, Becker admitted that he had been in doubt about titling his 1964 book *Human Capital* because the term "…was alleged to be demeaning because it treated people as machines" (Becker, 1993). However, Becker also pointed out that education offers non-pecuniary and non-market types of return (Alstadsæter, 2003). One of his students, Michael Grossman (1972a, 1972b), developed what is now known as the classical model of demand for health. In the model, health capital is seen as one component of the stock of human capital: being in good health yields utility in itself as well as income through market production. Health as an investment commodity determines the total amount of time available for market and non-market activities. Health is produced by means of the individual's use of his or her own time and services bought in the market. Thus, the demand for health services is in turn derived from the demand for health. In Grossman's model, the private return to investment in health may, broadly speaking, be measured by the number of illness-free days that an individual enjoys in any given year. Likewise, education that increases productivity will, in a perfect labour market, yield a return through higher wages.

In addition to the private return, however, investments in health and education can also yield a social return, i.e., to persons other than the one undertaking the investment. In the health domain, one person's lifestyle may bear consequences for other peoples' behaviour, e.g., smoking, eating habits, or level of activity. The social return to education is commonly associated with the diffusion of general knowledge, which makes other persons more productive (Lucas, 1988). Education may also have externalities in more subtle ways, e.g., through implications for the pattern of human fertility.

Like other forms of capital, human capital will depreciate. Education, skills and knowledge are forgotten or can become obsolete. Relating the human capital terminology to hip replacements, we can say that the demand for an operation is derived from the demand for health. Even from birth, people differ in their health stock: some patients can have a hip defect from when they are only a few months old (developmental *dysplasia* of the hip). Over time, the hip joint can also be damaged from long usage, so *arthritis* is the most common cause of hip replacement.

As pointed out by Kenneth Arrow (1963) in his seminal article, the health-care sector is characterized by a high degree of asymmetric information, e.g., the patient does not know which treatments are available for a particular illness and cannot easily compare the quality of health services offered. Arrow suggested that such asymmetries help explain why non-market health institutions arise. Several countries where health services are publicly financed have initiated competition in the health-care sector through patient choice of deliverer (Siciliani and Hurst, 2005). These reforms aim at improving efficiency by letting patients travel to institutions with idle capacity. The patients' motives for travel could be that the expected health improvement provides a higher quality of life, reflecting in turn the consumption aspect of health. Therefore, waiting for an operation bears a cost. Another cost of waiting could be the income foregone when the patient is excluded from the labour force due to illness. This reflects the human capital aspect. This brief overview presents two potential explanations for why education can be important for patient choice of hospital: the opportunity cost of time and information cost. Our analysis in chapter 1 investigates patients' preferences along several dimensions, one of which is education.

Economists have used human capital theory to explain the pattern of fertility. Hotz, Klerman and Willis (1997) give an overview of the literature. The basic idea is that taking care of children is time-intensive, and that the opportunity cost of time increases with education. As a result, more educated parents want fewer children, but may spend more resources on each child's education and upbringing (Becker, 1960; Willis, 1973). Gustafsson (2001) summarizes the theory on the timing of births and identifies the main factors as career planning and consumption smoothing. Gustafsson concludes that the main parameters that have an impact on career costs are the amount of pre-maternity human capital, the rate of depreciation of human capital from the non-use of human capital, the rate of return to human capital investments, the profile of human capital investments and the length of time spent out of the labour force. Chapters 2 and 3 elucidate upon the connection between education and fertility, analysed over the ages during which women are fertile. The outcome variables are the timing of first births and number of children, including childlessness.

A methodological problem when examining the link between education and fertility is how to identify the causal relationships. For instance, when the data show that the number of children decreases with education, is this because more educated parents wish to have fewer children because of the higher opportunity cost of time, or because individuals have different preferences that influence their choice of schooling as well as fertility? One way to overcome the identification problem is to employ "natural experiments", (see e.g., Angrist and Krueger, 2001). The fertility analysis in chapters 2 and 3 benefits from such a natural experiment: namely, an educational reform implemented in Norway from 1960 to 1972.

All articles analyse discrete choice, and a common feature is the use of a latent variable model where it is assumed that part of the utility derived from each alternative is observable to the researcher, and part is unobservable and treated as a random variable. Patient choice is estimated using a conditional logit model and fertility with a logit model. The fertility analysis is reduced-form estimation, while we use a structural model for the choice of hospital and estimate the marginal rate of substitution between distance and waiting time. Thus, in chapter 1, preferences are described, whereas in chapter 3, I examine factors that can shed light on how preferences are formed. Economists have become increasingly aware of the importance of the family as an institution for shaping values and habits. In this dissertation, teenage motherhood is analysed in terms of its relationship to schooling, as well as to family background and social interaction.

Summary of the chapters

The dissertation consists of three self-contained chapters. Chapter 1 makes use of a unique set of patient data originating from the Norwegian Arthroplasty Register and merged with data from the Norwegian Patient Register, Statistics Norway and a matrix of distances to investigate the impact of patient characteristics on the choice of hospital for elective care. Chapters 2 and 3 use a very rich data set of register data from Statistics Norway to analyse the causal determinants of fertility choices among Norwegian women, and the heterogeneity in their responses to educational reform and the effects of social interaction. The following provides a brief summary of each chapter.

Chapter 1: Patients' Preferences for Choice of Hospital
(Co-authors: Birgitte Espehaug and Lars Birger Engesæter)

Irrespective of the health system, patients' choice of hospital may be considered as the trade-off between price, distance and quality. In a national health system (NHS) where hospital treatment is close to free of charge at the point of treatment, price is irrelevant to the patient, but waiting lists typically occur (Cullis, Jones and Propper, 2000). These have been given considerable political attention. In fact, waiting time is one aspect of quality that is highlighted in health policy in several OECD countries. One of the supply-side policies used to reduce waiting time is to increase patient choice, and thereby enhance the competitive pressures on providers (Siciliani and Hurst, 2005). A recent ruling in the European Court of Justice extends patients' legal rights of choice dramatically within the European Union, as it gives patients within a NHS the option of publicly funded treatment abroad if they face any undue delay. In Norway, a reform launched in 2001 established a quasi-market for elective hospital care with the aim of equalizing waiting times across the country and improving capacity utilization. However, will paving the way for "market forces" in the hospital sector make any difference? To what extent a European or a national health market will emerge, depends, among other things, on patients' willingness to travel to reduce waiting time. As the willingness to pay for shorter waits may rarely be observed in the market, it must be inferred from actual behaviour or from surveys (Cullis *et al.*, 2000).

The contribution of this paper is to empirically analyse quality competition, focusing on the demand side and, more specifically, the trade-off between waiting time and distance. This trade-off is likely to differ between patient groups, and it should be easier to interpret the results when we focus on only a single patient group. In our analysis, patients' preferences are derived from their actual behaviour within a national health system, using a unique set of register data with individual patient information on socio-economic variables as well as medical data. Patient choice is analysed within a random utility framework using a conditional logit model.

We examine patients' preferences using data from 2001 to 2003 on patients undergoing primary total hip replacement (Furnes *et al.*, 2003). This is an interesting patient group for several reasons. First, hospital choice is an option for elective cases only, of which hip replacements constitute a large share (Christensen and Hem, 2004). Second, waiting times for this sort of treatment were substantial when the free choice reform was introduced: on average thirty weeks at a national level, notwithstanding large geographical variation. Third, the procedure is offered at many hospitals across the country.

The average age of the patient group is high, nearly 67 years. Quality differences among hospitals have been detected, as the risk of revision is found to be less in hospitals where surgeons perform a high number of operations each year (Espehaug *et al.*, 1999; Losina *et al.*, 2004). Because total hip replacement is a quite common type of surgery, we would expect general practitioners (GPs) to have a general opinion on the quality of different hospitals. The fact that information on prostheses survival related to individual hospitals or surgeons is not published in Norway should not rule out competition based on general reputation or observable quality aspects such as waiting time.

A general finding in the literature on hospital choice is that distance is important. Tay (2003) refers to studies that identify various proxies for hospital quality: capacity, high volume, the range of services, the complication rate, the mortality rate, etc. For hip replacements specifically, the quality criterion most often used in the medical literature is survival of the prosthesis. In this study, we assume that quality aspects other than waiting time are captured by a set of hospital dummies. These dummies represent various dimensions of perceived quality that are fixed within the study period, and in principle observable both to the patient and to the researcher, but not included separately in the analysis, e.g., university hospital status or general reputation.

We find that distance is a very important attribute when patients consider hospital choice for elective hip replacement. Waiting time is also estimated to be statistically significant and to have a negative effect on utility, but its impact on behaviour is found to be small. Given the marginal effect of waiting time on utility is found to be negative rules out the possibility that long waiting lists can be regarded as a signal of good quality. The model includes a hospital-specific fixed effect, which should cover time-constant effects, such as reputation.

The estimated trade-off between distance and waiting time varies considerably between models and patient categories. Patients are categorized according to age, gender, education and the year of referral. Avoiding distance is especially important to older patients, and the estimates show no statistically significant gender differences. Clearly, the most important factor for the estimated marginal rate of substitution is the level of education. Irrespective of age, gender and the year of referral, a patient with more education is more willing to travel and less willing to wait. In the estimated sample, the mean patient in each category is less reluctant to travel for an operation in 2003 than in 2001, although this result is not robust to changes in sample size.

The most striking finding is the great reluctance to travel among patients having a primary hip replacement. The most mobility-inclined patient (as measured by the marginal rate of substitution), represented by a man under the age of 67 years with higher education who entered the waiting list in 2003 must, on average, benefit from a reduction in waiting time of 32 weeks to be willing to travel just one extra hour.

Chapter 2: Education and Fertility: Evidence from a Natural Experiment (Co-authors: Carol Propper and Kjell G. Salvanes)

Fertility continues to be an issue of public concern, even in developed countries that have experienced the demographic transition and reached a state where both mortality and birth rates are low. Low population growth and higher dependency ratios are argued to strangle economic growth. Recent OECD projections suggest that, because of demographic changes, the growth rate of per capita income will decline from 1.7% to 1.1% by 2050 in European countries and from 1.7% to 1.2% in the United States (Turner *et al.*, 1998). Often when low birth rates and fertility patterns are discussed, women's trade-off between childcare and education and employment opportunities are brought forward as one explanation. The observed relationship between fertility and female education varies between different countries and time periods, but there is much empirical support for strong correlations (Schultz, 1997; Cochrane, 1979). However, many factors influence decisions on fertility, education and employment, very likely including unobservable factors that cannot be controlled for. Thus, causation is difficult to establish. In this paper, we make use of an educational reform to trace the causal effect of education on fertility outcomes.

Nordic countries have a relatively high fertility rate (Sleebos, 2003), but this is an imperfect measure of long-run fertility as it aggregates behaviour over cohorts and ignores the timing of births. With respect to population development that is sustainable, the major concern in Nordic countries is the increasing number of childless women and the fact that the younger cohorts of women are having fewer children (Skrede and Rønsen, 2006). Our data enables us to estimate the effect of education on the timing of births as well as completed fertility, including the probability of being childless, after allowing for cohort effects. As the cohorts studied were born between 1946 and 1958, our data includes the most recent generation of women with completed fertility histories.

We study the relationship between the education of women and three fertility outcomes: the timing of children; childlessness; and the number of children. Our data confirms the expected correlation between fertility outcomes and education: women with more education are more often childless; they have fewer children and postpone births. Despite these statistically significant correlations, we do not find evidence of a causal relationship between the length of education on one hand, and completed fertility or childlessness on the other, when using the reform as an instrument for education. Our main finding is that increased mandatory education lead to the postponement of births; there are fewer cases of teenage motherhood and more first births among women aged 35 to 40 years. This result cannot be explained as a mere "incarceration effect", and we interpret it mainly as a result of increased human capital accumulation from the reform.

Chapter 3: Education and Fertility: Testing for Family Background and Spillover Effect

Studying the causal relationship between fertility and education, Monstad, Propper and Salvanes (2007) find that more education leads women to postpone first births, but that it does not result in lower total fertility or the greater incidence of childlessness. The causality is based on a natural experiment, i.e., an educational reform that increased compulsory schooling in Norway by two years. The effect estimated is by definition a "local average treatment effect" (Angrist, 2004). This naturally raises questions about the generality of the results. Policy measures are often intended to benefit certain segments of the population, which is another reason to study heterogeneity in policy response. Indeed, one of the main aims of the educational reform in question, as stated explicitly in government documents, was to enhance the equality of opportunity along both socio-economic and geographic dimensions (Black, Devereux and Salvanes, 2005a). Furthermore, if education has a causal impact on fertility, particularly the timing of births, this is a potential channel through which education can have distributional consequences across generations.

Investments in education can be evaluated by the private rate of return. If externalities arise, the social and private rate of return will differ (Lucas, 1988). Even if educational reforms are hardly ever implemented because of their effect on fertility, one should bear in mind that such policy measures have fertility consequences and that fertility behaviour implies externalities. For instance, at the macro level, the number of children born and the age structure of the population have implications for economic growth. Research also suggests that teenage pregnancy shapes the life conditions for the child to be born in an adverse manner (for references, see Black *et al.*, 2006). Moreover, motherhood at a later age also can have unfavourable medical consequences for the child: "…more stillbirths, more infant deaths, more premature births, more chromosomatic problems and more learning problems" (Gustafsson, 2001, p. 244).

One way that externalities can arise is that one person's behaviour and norms may shape another person's preferences and behaviour. Such spillover effects are a special concern in the "new social economics literature" (Durlauf and Young, 2001). This literature examines such diverse phenomena as residential segregation (Schelling, 1971), neighbourhood effects on teenage childbearing (Crane, 1991) and how the presence of other smokers in a household affects the decision to quit smoking (Jones, 1994). Fertility is influenced by many factors, e.g., economic and cultural factors. It then appears reasonable that the family is an institution that shapes young girls' values and attitudes towards important decisions, including the choice of education and family formation. In several studies, the characteristics of the family have proven to have a great impact on young people's choice of education, labour market outcomes, etc. (see e.g., Aakvik, Salvanes and Vaage, 2005; Black *et al*., 2005a and 2005b; Raaum, Salvanes and Sørensen, 2006). In this paper, I examine whether community and family background also play an important role in decisions on fertility, and whether a spillover effect can be traced in the data. Elder relatives (grandparents, uncles and aunts) have been proven to have an impact on educational outcomes for same-gender adolescents (Loury, 2006). I estimate the impact on fertility of elder sisters' education, while also controlling for the mother's and father's education.

The purpose of this paper is twofold. First, to examine the extent to which there is heterogeneity in the response to educational reform, and thereby identify the groups of women whose fertility behaviour changed due to the reform. Second, to examine whether education triggers a spillover effect within the family, so that an elder sister's having more compulsory education has an impact on the younger sister's fertility outcomes, in particular, the probability of teenage motherhood. Moffitt (2001) points to several methodological problems in identifying the effect of social interactions. This analysis benefits from a natural experiment, this help solve the problem of unobservable heterogeneity. Unlike many other studies, the impact of family background is studied within the context where the link between education and fertility is causal.

Family background proves to be an important causal determinant for fertility behaviour in general, but also for the effect of educational reform on fertility. The analysis shows much heterogeneity in the response to educational policy. In particular, the effect depends on family income and whether the young woman lives in a city. The heterogeneity in the response is especially strong regarding the likelihood of first birth as a teenager. The group that responded to the reform most strongly in terms of delaying first birth consists of women from low-income families, living in cities. These women also show an increase in the tendency to remain childless. However, the effect of family background does not seem to incorporate spillover effects of the reform from elder to younger sisters within the same family. The spillover effect of the reform is estimated to have the expected sign (to reduce teenage motherhood), but it is of small magnitude and statistically insignificant.

Regarding the intention to enhance the equality of opportunity, it is worth noting that as a consequence of the reform, the timing of first births and especially the frequency of teenage motherhood has become more similar among the different income groups. Along the urban/non-urban dimension, the picture is more mixed. Using a specification that focuses on the poorest income quartile, I find that the gap between urban and non-urban women is diminished because of the reform.

References

Aakvik, A., K.G. Salvanes and K. Vaage (2005). "Educational attainment and family background", *German Economic Review*, vol. 6(3), pp. 377–394.

Alstadsæter, A. (2003). *"Income Tax, Consumption Value of Education and the Choice of Educational Type"*, The Norwegian School of Economics and Business Administration, Discussion Paper 11/2003.

Angrist, J.D. (2004). "Treatment effect heterogeneity in theory and practice", *The Economic Journal*, vol. 114, pp. 52–83.

Angrist, J.D. and A.B. Krueger (2001). "Instrumental variables and the search for identification: From supply and demand to natural experiments", *Journal of Economic Perspectives*, vol. 15(4), pp. 69–85.

Arrow, K. (1963). "Uncertainty and the welfare economics of medical care", *The American Economic Review*, vol. 53, pp. 941–973.

Becker, G. (1960). "An economic analysis of fertility", in: *Demographic and Economic Change in Developed Countries*, Universities-National Bureau of Economic Research Conference Series 11 (NBER), pp. 209–231.

Becker, G. (1964). *Human Capital*. 2nd edn, New York: Columbia University Press, 1975.

Becker, G. (1993). "Nobel Lecture: The economic way of looking at behavior", *The Journal of Political Economy*, vol. 101(3), pp. 385–409.

Black, S., P. Devereux and K.G. Salvanes (2005a). "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital", *American Economic Review*, March, vol. 95(1), pp. 437–449.

Black, S., P. Devereux and K.G. Salvanes (2005b). "The more the merrier? The effect of family composition on children's outcomes", *Quarterly Journal of Economics*, May, vol. 120(2), pp. 669–700.

Black, S., P. Devereux and K.G. Salvanes (2006). "Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births", NBER Working Paper No. w10911. Forthcoming in *The Economic Journal*.

Christensen, Ø. and K.G. Hem (2004). "Fritt sykehusvalg i Norge", SINTEF Helse, report no. STF78 A015010 (in Norwegian). ["Free choice of hospital in Norway"].

Cochrane, S.H. (1979). *Education and Fertility: What Do We Really Know?* John Hopkins University Press, Baltimore.

Crane, J. (1991). "The epidemic theory on ghettos and neighborhood effects on dropping out and teenage childbearing", *American Journal of Sociology*, vol. 96, pp. 1226–1259.

Cullis, J., P.R. Jones and C. Propper (2000). "Waiting lists and medical treatment: analysis and policies", ch. 23 in A.J. Culyer and J.P. Newhouse (eds), *Handbook of Health Economics*, Elsevier.

Durlauf, S.N. and H. Peyton Young (2001). "The New Social Economics", ch. 1 in *Social Dynamics*, S.N. Durlauf and H. Peyton Young (ed), Brookings Institution Press, Washington, D.C.

Espehaug, B., L.I. Havelin, L.B. Engesæter and S.E. Vollset (1999). "The effect of hospital-type and volume on the survival of total hip replacements. A review of 39505 primary hip replacements reported to the Norwegian Arthroplasty Register, 1988–1996", *Acta Orthop Scand.*, vol. 70(1), pp. 12–18.

Furnes, O., L.I. Havelin, B. Espehaug, L.B. Engesæter, S.A. Lie and S.E. Vollset (2003). "Det norske leddproteseregisteret—15 nyttige år for pasientene og for helsevesenet", *Tidsskrift for den Norske Lægeforening*, vol. 123(10), pp. 1367–1369 (in Norwegian)

["The Norwegian Arthroplasty Register—15 years of service to the patients and the health care system"].

Grossman, M. (1972a). "*The Demand for Health: A Theoretical and Empirical Investigation"* (New York: Columbia, for the National Bureau of Economic Research).

Grossman, M. (1972b). "On the concept of health capital and the demand for health", *Journal of Political Economy,* vol. 80: 223–255.

Gustafsson, S., (2001). "Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe", *Journal of Population Economics,* vol. 14(2), pp. 225–247.

Hotz, V.J., J.A. Klerman and R.J. Willis (1997). "The Economics of Fertility in Developed Countries" , ch. 7 in Rosenzweig, M.R., Stark, O. (eds), *Handbook of Population and Family Economics*, vol. 1A. Elsevier Science, North-Holland, Amsterdam, pp. 276-348.

Jones, A. (1994). "Health, addiction, social interaction and the decision to quit smoking", *Journal of Health Economics,* vol. 13, pp. 93–110.

Losina, E., J. Barrett, N.N. Mahomed, J.A. Baron, and J.N. Katz (2004). "Early Failures of total hip replacement effect of surgeon volume", *Arthritis & Rheumatism,* vol. 50(4), pp. 1338–1343.

Loury, L.D. (2006). "All in the extended family: Effects of grandparents, aunts and uncles on educational attainment", *American Economic Review Papers and Proceedings*, vol. 96, pp. 275–278.

Lucas, R. (1988). "On the mechanics of economic development", *Journal of Monetary Economics,* vol. 22 (July), pp. 3–42.

Moffitt, R. (2001). "Policy interventions, low-level equilibria and social interactions", ch. 3 in *Social Dynamics*, S.N. Durlauf and H. Peyton Young (eds), Brookings Institution Press, Washington, D.C.

Monstad, K., C. Propper and K.G. Salvanes (2007). "Education and fertility: Evidence from a natural experiment", chapter 2 in Doctoral thesis, the Norwegian School of Economics and Business Administration.

Raaum, O., K.G. Salvanes and E.Ø. Sørensen (2006). "The neighbourhood is not what it used to be", *The Economic Journal,* vol. 116 (January), pp. 200–222.

Schelling, T. (1971). "Dynamic models of segregation", *Journal of Mathematical Sociology,* vol. 1, pp. 143–186.

Siciliani, L. and J. Hurst (2005). "Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries", *Health Policy,* vol. 72(2), pp. 201–215.

Sleebos, J. (2003). *Low Fertility Rates in the OECD Countries: Facts and Policy,* Working Paper No. 15, Directorate for Employment, Labour and Social Affairs, OECD Paris.

Skrede, K. and M. Rønsen (2006). "Hvor bærekraftig er den nordiske familiepolitiske modellen?", *Samfunnsspeilet,* vol. 19(2), pp.2-10 (in Norwegian only) [How sustainable is the Nordic family policy model?"].

Tay, A. (2003). "Assessing competition in hospital care markets: the importance of accounting for quality differentiation", *Rand Journal of Economics,* vol. 34(4), pp. 786–814.

Willis, R. (1973). "A new approach to the economic theory of fertility behavior", *Journal of Political Economy,* vol. 81(2), pp. S14–S64.

Chapter 1

Patients' Preferences for Choice of Hospital*

by

Karin Monstad
Department of Economics
Norwegian School of Economics
karin.monstad@nhh.no

Lars Birger Engesæter
The Norwegian Arthroplasty Register
Department of Orthopaedic Surgery, Haukeland University Hospital

Birgitte Espehaug
The Norwegian Arthroplasty Register

***Authors' Declaration:**
Karin Monstad has the sole responsibility for all economic and econometric analysis in this article. Espehaug's and Engesæter's roles have been to make data from *The Norwegian Arthroplasty Register* available and to assist with the interpretation of these data from a medical perspective.

Acknowledgements:
We are grateful for valuable comments from Fred Schroyen, Jan Erik Askildsen, Carol Propper, Kjell Gunnar Salvanes, Andrew Jones, Marten Lindeboom, Kurt Brekke, Luigi Siciliani and participants at the Health Economics Workshop in Leuven, April 2006 and for great help with the data set and the econometric analysis from Tor Helge Holmås and Erik Ø. Sørensen.

Abstract

What determines patients' choice of hospital, in a setting where hospital stays are rationed by waiting lists and where travel distances within the country are substantial? Through a reform implemented in 2001, Norwegian patients are given generous formal rights to choose any hospital throughout the country for elective treatment. This paper is an attempt to infer the willingness to pay for shorter waits by studying the observed allocation of operations. The trade-off between distance and quality is likely to differ according to patient characteristics. Patients' preferences are examined using a unique data set with individual patient data on one specific patient group, namely elective total hip replacements in Norway during the years 2001–2003. After a discussion of the institutional setting, the paper focuses on the trade-off that the patients make between distance and waiting time, and explores whether quality competition can be traced in the Norwegian hospital sector. The main results are that distance and waiting time are both highly statistically significant attributes, and that patients are willing to wait a considerable length of time to avoid travelling. The reluctance to travel is found to increase with age and decrease over time and with the level of education.

JEL classification: I11, C25, D12

Key words: hospital choice, waiting times, elective surgery, competition.

1.1 Introduction

Irrespective of the health system, patients' choice of hospital may be summed up as a trade-off between price, distance and quality. In a national health system (NHS) where hospital treatment is close to free at the point of treatment, price is irrelevant to the patient, but waiting lists typically occur (Cullis *et al.*, 2000) and have been given considerable political attention. In fact, waiting time has been the one aspect of quality that is highlighted in health policy in several OECD countries. One of the supply-side policies used to reduce waiting time is to increase patient choice and thereby enhance competitive pressures on providers (Siciliani and Hurst, 2005). A recent ruling in the European Court of Justice extends patients' legal rights of choice dramatically within the European Union, as it gives patients within an NHS the option of a publicly funded treatment abroad if they face undue delay.[1] In Norway a reform was launched in 2001, which established a quasi-market between hospitals with the aim to equalize waiting times across the country and improve capacity utilization.[2] However, will paving the way for "market forces" in the hospital sector make any difference?[3] To what extent a European or a national health market will emerge, depends, among other things, on patients' willingness to travel to reduce waiting time. As the willingness to pay for shorter waits may rarely be observed in the market, it must be inferred from actual behaviour or from surveys (Cullis *et al.*, 2000). The contribution of this paper is to analyse quality competition empirically, focusing on the trade-off between waiting time and distance. Patients' preferences are derived from their actual behaviour within a national health system, using register data with information on patient heterogeneity.

Patients' preferences are examined using data from 2001 to 2003 on a specific patient group, namely patients with primary total hip replacements (Furnes *et al.*, 2003). (See the appendix.) The empirical work uses a unique data set with individual patient information on socio-economic variables as well as medical data. The focus is on the demand side, and the starting point of the analysis is that all patient movement within this particular patient group is

---

[1] The ruling concerned the case of Yvonne Watts, a 75-year-old British woman who claimed compensation from her Primary Care Trust after she paid to have a hip operation in France (www.news.bbc.co.uk and www.curia.eu.int). The legal rights seem to be the same as are already implemented in Norway (as of the 1st of September, 2004), but may cause changes in EU member states where services are rationed by waiting times, e.g., the UK.

[2] Hoel and Saether (2003) present arguments why a reduction in waiting times for public health treatment may not be welfare increasing.

[3] In his "Letter from America", Angus Deaton (2006) has given a vivid description of the problems of getting good information on quality and price, based on his own experience as a hip replacement consumer.

to be regarded as a choice that reflects patients' preferences, given the information they have. Of course, we only observe the actual behaviour, i.e., where the operation took place and the wait experienced. The alternatives actively considered by the different parties (patient, GP and hospital) are not known. However, patients' alternatives are described by available information on travel distances and average waiting time at different hospitals.

A general finding in the literature on hospital choice is that distance is important. Tay (2003) refers to studies that identify various proxies for hospital quality: capacity, high volume, range of services, complication rate, mortality rate etc. For hip replacements specifically, the quality criterion most often used in the medical literature is survival of the prosthesis (see the appendix). In this study, we assume that quality aspects other than waiting time are captured by a set of hospital dummies. These dummies represent dimensions of perceived quality that are fixed within the study period and in principle observable both to the patient and the researcher, but not included separately in the analysis, e.g., university hospital status or general reputation.

This patient group is interesting for several reasons. Hospital choice is an option for elective cases only, of which hip replacements constitute a large patient group (Christensen and Hem, 2004). Waiting times for hip replacements were substantial when the free choice reform was introduced, on average 30 weeks at a national level, with great geographical variation. The procedure is offered at many hospitals across the country. The average age of the patient group is high, nearly 67 years. Quality differences among hospitals have been detected, as the risk of revision is found to be less in hospitals where surgeons perform a high number of operations per year (Espehaug *et al.*, 1999; Losina *et al.*, 2004). Because total hip replacement is a type of surgery that is quite common, we would expect GPs to have a general opinion on the quality of different hospitals. The fact that information on prostheses survival related to individual hospitals or surgeons is not published in Norway should not rule out competition based on general reputation or observable quality aspects like waiting time.[4]

The trade-off between distance and quality is likely to differ between patient groups. It should be easier to interpret the results when we, like Tay, focus on only one patient group. Vrangbæk *et al.* (2006) provide an overview of the evidence about patients' awareness of the right to choose a hospital, and the data on patient movement in the Scandinavian countries. This paper is an attempt to add new insight by studying the revealed preferences of individuals within a specific patient group, also using data on socio-economic background.

[4] For more information on quality aspects of hip replacements, see the appendix.

Patients' choice is analysed within a random utility framework, using a conditional logit model.

The structure of the paper is as follows. Before the theoretical framework and the hypotheses are presented in section 3, the institutional framework is explained in some detail in section 2. Data are described in section 4, and section 5 explains the empirical specification used. The estimation results are presented and discussed in section 6. Section 7 concludes.

1.2 Institutional framework

Several European countries have introduced policies to enhance choice in health care (Siciliani and Hurst, 2005). Vrangbæk *et al.* (2006) point out that "[t]he Nordic experience presents a unique opportunity to study patients' choice and the hospitals' reactions to choice in a situation with little or no interference from user payments, no incentives for the GPs to refer to certain hospitals, and strong economic incentives for the hospitals to attract patients". In the setting described, we find it valid to study patient movement by focusing on patient characteristics, interpreting their behaviour as an expression of their preferences and implicit costs. In the following, we shall outline the institutional framework in more detail.

1.2.1 Demand-side incentives and restrictions

Norway's health system is largely financed by general taxes. Most services are nearly free of charge at the point of usage. Norwegian patients have been granted a legal right to choose a provider for elective treatments in somatic or psychiatric specialist care, whether as an inpatient or outpatient.[5] The Patients' Rights Act was implemented on the 1st of January, 2001. Patients' co-payment for transportation is in most cases negligible, about 27 Euros (220 Norwegian Kroners (NOK)) one way if the patient goes to a hospital in another health region, about 16 Euros (115 NOK) otherwise (payment data are for 2005).

For a large part of the population, sickness allowance is 100 per cent of the patient's regular wage during the first year of sickness leave.[6]

[5] Patients cannot require to be treated at a more specialized institution than the one he or she was referred to, but this restriction is not binding, because all Norwegian hospitals also function as local hospitals (Christensen and Hem, 2004). The right extends to all public hospitals in the country. It was taken as granted that "public hospitals" included private non-commercial hospitals that had an agreement with hospital authorities (Ot.prp. no. 63 (2002–2003)). The patient choice was extended to private commercial hospitals by the 1st of September, 2004, which is outside the scope of this study.

[6] Self-employed and employees with high income are not automatically fully insured through the National Social Security System.

The patient is usually referred to a hospital by a GP. To assess whether a hip replacement is necessary, there is typically an examination by an orthopaedic surgeon at an outpatient clinic. The referral implies that the patient is placed on a waiting list at a particular hospital. The patient may switch to another hospital while waiting, but will then be treated as a newcomer to the latter hospital's waiting list, so there is a certain lock-in. Waiting time is defined as the time elapsed between referral and the date of hospitalization.

Information on waiting times has been made available at a free telephone service starting when the reform was implemented in 2001. More than 20000 persons called this number in 2003 (Godager and Iversen, 2004).[7]

1.2.2 GP's incentives

Whether it is the patient or the GP who makes the choice of hospital is important if the medical advisor has other preferences and/or possesses other information than the patient. The GP is likely to be better informed about the overall quality of different hospitals. Through a reform introduced June 1, 2001, each Norwegian citizen is entitled to a specified GP who is given a key role as advisor when patients choose a hospital. Most GPs are self-employed and they are financed partly by list patient capitation and partly by fee-for-service. The GP himself has no economic incentives to refer to specific hospitals. Gathering information is time-consuming and therefore costly to him (Vrangbæk *et al.*, 2006). The GP gets no direct compensation for such services, but the competition for patients introduced by a list-capitation system may give stronger incentives to engage in the matter (Carlsen *et al.*, 2005). Even if one is not willing to regard the GP as a perfect agent for the patient in general (McGuire, 2000), it is difficult to see what self-interest a GP should have in making referrals to a specific hospital, except for possible loyalty and personal relations. Still, patients may differ in their search cost. If the GP does not engage in giving information on hospital choice, differences in patients' search costs may be decisive for observed patient behaviour.

1.2.3 Hospital incentives

Total hip replacements are carried out by the majority of Norwegian hospitals, but the number of operations per year varies significantly among them.

The government allocates its budget to health regions, which are free to decide on what basis individual hospitals under their jurisdiction should be remunerated.[8] Since 1997,

[7] In May 2003 the Government launched an information service on the Internet, www.sykehusvalg.no. This study uses data for patients who entered the waiting list no later than June 2003.

hospital owners have been given economic incentives to attract patients, as part of their remuneration has been based on activity level. The rest is given as a block grant. The part that is paid based on activity was 50% of the stipulated cost per diagnosis-related group (DRG) in 2000 and 2001, 55% in 2002 and 60% in 2003 (BUS, 2005). For patients who cross health regions, the payment must be settled in an agreement between the two health regions involved. If no agreement is made, there is a standard norm stipulated by the Ministry of Health. The standard norm is 80% of the stipulated DRG cost.

There has been some publicity on allegations that hospitals specialize in some well-paid treatments (e.g., snoring operations) because payment compared to costs varies significantly both between and within DRGs. Until 2003, all hip replacements were defined in one category, DRG 209, with a stipulated cost of about 13,700 Euros. In 2003 a subcategory for complicated cases was introduced, DRG 209B, for which the compensation per treatment was about 2,000 Euros higher. Elective surgery, including hip replacements, is considered to be an economically and organizationally attractive activity for an orthopaedics department.[9]

The costs of transportation of patients in specialized care did not affect local or regional health authorities in the period studied.[10]

Hospitals that are affected by the reform have a duty to “accept all patients who choose the hospital” (Ot.prp. no 63, 2002–2003) but have a formal right to reject patients from another health region if they need to prioritize their own patients for capacity reasons (Directorate for Health and Social Affairs, circular IS-12/2004).

1.3 Theoretical framework and hypotheses

The basic notion is that patients have preferences over different attributes of hospital treatment. Relevant attributes could be travel cost, waiting time, post-operative mortality, complication rate, and survival of the prosthesis. Patient i is assumed to choose a hospital $h = (1,..,H)$ so as to maximize the utility function:

$$U_i(D_{ih}, W_h, q_h, Z_{ih}), \qquad (1)$$

[8] In 2000 and 2001, public hospitals were owned by 19 different counties. By the hospital reform implemented Jan.1, 2002, the country was divided into five Regional Health Authorities who themselves own “hospital enterprises”, which own individual hospitals.
[9] According to an internal report from one of the Regional Health Authorities (also called “Health Regions”), elective orthopaedics is profitable to the orthopaedics department. To have a high volume of operations gives status and attracts candidates for specialization (Helse Nord, 2003)
[10] By January 1, 2004 the financial responsibility for transportation costs was placed with the regional health authorities, to give incentives so that the patient is treated near his home “when this is beneficial to the patient and reduces the cost of transportation” (Department of Health, 2005).

where D is distance to hospital, W is waiting time, q is a vector of other observable quality attributes, and Z is quality that is known to the demander, but not observed by the researcher. We focus on two of the elements of U_i, namely D and W, and expect that $\frac{\delta U}{\delta D} < 0$, $\frac{\delta U}{\delta W} < 0$, and by appropriate choice of units, that $\frac{\delta U}{\delta q} > 0$ and $\frac{\delta U}{\delta Z} > 0$.

Receiving treatment adds to utility because of health improvement, so there is an opportunity cost to staying on the waiting list. The purely health-related waiting cost may consist of several elements: foregone expected benefit, which depends on discounting, temporary pain while waiting and possibly a higher risk of a permanent reduction in health status (Siciliani, 2005). Whether waiting also results in a monetary loss depends on how well the patient is insured. As the expected average waiting time W_h differs between hospitals, so does the waiting cost. Note that the waiting time at hospital h is assumed to be the same for all patients. This could be because the patient is only informed about the *average* expected waiting time and is not given an individual expected waiting time at hospital h[11], or because there is no prioritization according to need nor any cream-skimming taking place.

There are also some costs attached to receiving treatment. The disutility connected to specific procedures executed at the hospital is considered equal for all hospitals. What may differ between hospitals is the patient's perceived *travel costs*. These costs are to be considered mainly non-monetary, reflecting the unease of travelling long distances and being away from relatives and friends during the hospital stay.[12] They depend on the patient's preferences and the travel distance or time, D_{ih}.

Given (1), we can describe a utility-maximizing patient's trade-off between D and W using the marginal rate of substitution: $MRS_i \equiv -\frac{dW}{dD}|_{dU=0}$.

The possibility that the patient will not have the operation at all is represented by the alternative $(D_{i0}, W_0, q_0, Z_{i0})$, which is the outcome if travel distance and waiting time are very high, or if other quality elements are very poor. The patient therefore faces an opportunity set A_i, where $A_i = \{(D_{ih}, W_h, q_h, Z_{ih})_{(h=1,..H)}, (D_{i0}, W_0, q_0, Z_{i0})\}$.

[11] However, by a law enforced on 1st September, 2004 all patients having elective operations are entitled to an individually set waiting time.
[12] The average length of stay at hospital is about 11 days for hip replacements. The possibility that the patient regards travelling to certain perhaps distant destinations as a good rather than a bad is ruled out, although it is conceivable. See ww.aftenposten.no/forbruker/helse/article848076.ece

The patient's problem is to maximize (1) with respect to *h*, subject to $(D_{ih}, W_h, q_h, Z_{ih}) \in A_i$. If hospital *j* is chosen by *i*, then:

$$U_i(D_{ij}, W_j, q_j, Z_{ij}) \geq U_i(D_{ih}, W_h, q_h, Z_{ih}), h = 0,....,H .$$

For simplicity, utility is assumed to be an additively separable function in the arguments and also to be linear in *q* and *Z*, so that for any given patient:

$$U_{ih} = f(D_{ih}, X_i; \alpha) + g(W_h, X_i; \beta) + \gamma q_h + Z_{ih}, \qquad (2)$$

where *f*(.) and *g*(.) allow distance and waiting time to enter non-linearly, X_i is a vector describing patient *i*'s characteristics, and α, β and γ are parameter vectors. The *f* and *g* functions and the parameters are to be specified in greater detail in section 5. The patient's choice of hospital is discrete and may be illustrated as shown in figure 1 (the figure is drawn for convex preferences, but non-convexity is also conceivable).

In Figure 1, the patient prefers hospital A to hospital B, because a shorter waiting time more than compensates for the extra travel. However, a corner solution with $D_{ih} = 0$ is the best attainable, so the closest hospital, C, is chosen even though it offers a much longer waiting time than A. Judged by the two attributes D_{ih} and W_h, hospital D is the best alternative. The model implies that if D is not chosen, it is because D scores poorly compared to C on Z_{ih} or q_h. Finally, we assume that hospitals want to attract as many patients as possible, which is consistent with profit-maximizing hospitals receiving a payment per treatment that exceeds marginal cost for all *i*'s.

1.3.1 Hypotheses to be tested

The hypotheses that we want to test are the following:

1. Main hypothesis: patients dislike both waiting and travelling for an operation. They may be willing to travel to a more distant hospital if they are compensated through shorter expected waiting times.
2. There should be significant differences between those who travel and those who do not on observable characteristics that according to theory influence subjective waiting costs and travel costs.

The first hypothesis states that indifference curves are negatively sloped in the (D,W) space, although one cannot rule out that patients dislike a very short waiting time because they may want time to make arrangements before having the operation. The second hypothesis says that the marginal willingness to pay for a reduction in waiting time depends on socio-economic characteristics. (We shall define MRS as the reduction in waiting time needed for patients to be willing to travel to a more remote hospital.) For example, older people should be less willing to travel because their travel costs are higher; they are frailer in general, controlling for diagnosis. They may also be less able to gather information on waiting times, while we expect education to lower information search costs. The effect of gender is difficult to predict. There should be no effect through the labour market, if patients are fully insured. Any indirect gender effect through parenthood is difficult to measure in the sample, and its expected direction is also unclear.[13]

We shall also examine whether patients' behaviour has changed over time. One might expect that over the years, as information about the reform was more widespread, patients would reveal more reluctance to wait and less reluctance to travel. Being informed about patients' rights is a necessary condition for patients to choose a hospital further away. It is not a sufficient condition, because even well-informed patients may prefer to have the operation close to their home. Therefore we cannot use data on observed behaviour to infer whether patients' access to information has improved over time. We still find it interesting to examine whether it is possible to trace a year effect, even if the interpretation is not clear.

1.4 Data

The data set is a pooled cross-section obtained by merging data from four different sources. Details on these data sets and the exclusion criteria follow below. The source data are from *The Norwegian Arthroplasty Register* and this paper uses data on primary hip replacement operations performed during the period 2001–2003. The data set for analysis consists of 9753 observations/patients, who lived in 427 of Norway's 434 municipalities. The operations took place at 62 hospitals distributed in 55 different municipalities. The patients' choice set is the same during the period except that one hospital did not operate in 2003 and another one is only present in the 2002 data.[14] For each operation there is information on patient

[13] In this sample, only 14% of the patients had children under the age of 18 years. A gender effect via parenthood requires an assumption that the parent role means more to women than to men, and that parenthood influences preferences in a certain direction. Having (young) children may impose higher waiting costs, e.g., from not being able to participate in activities. On the other hand, being far away from children causes travel costs to rise.

[14] Of the patients on the waiting list, 3866 entered the list in 2001 and 1917 in 2003. The total number of observations is therefore (9753*62) – 3866 – (1917*2) = 596986.

characteristics and hospital characteristics for each possible choice that the patient could make.

1.4.1 Descriptive statistics

The variables used are described in Table 1, which also shows some other variables that may be of interest. The dependent variable takes the value 1 if individual i has chosen hospital j, and 0 if individual i has chosen $h \neq j$. Key hospital characteristics are expected waiting time (in weeks) and travel time by car (in hours) from the patient's home municipality to the hospital municipality. The expected waiting time at hospital h in year t is set equal to the mean actual wait at hospital h in year t, where t refers to the year when the patient was registered on the waiting list. Important patient characteristics are gender, age at referral, level of education and the year the patient was placed at the list.

The reference individual is a man under the age of 67, who entered the waiting list in 2001, with less than completed secondary education. Seventy per cent of the patients are women and the average age is nearly 70 years. Thirty-nine per cent entered the list in 2001, 41% in 2002 and, because of truncation of the data, about 20% in 2003. Twenty-five per cent of the patients had completed at least secondary education.[15] For the *alternatives actually chosen*, the average expected waiting time is 22,4 weeks and the average travel time is 1,1 hours. The mean values for *all possible* choices that a patient could make are 24,1 weeks and 11,5 hours, respectively. Forty-one per cent of the patients had the operation at a hospital other than the closest one that offers hip replacements.

1.4.2 Construction of the data set

The main data set is from *The Norwegian Arthroplasty Register* (hereafter NAR) and consists of operations done during the period 2000–2003. Registrations are voluntary and based on registration forms that the surgeon fills in right after the operation. Both public and private hospitals report to the register, which in recent years has had a reporting rate of 98% of all hip replacements (Espehaug *et al.*, 2006). The file registers 28862 operations on 25607 individuals. For the purpose of this paper, only primary hip replacements and treatment at Norwegian hospitals were considered, so 24925 observations are relevant.[16] NAR has data on

[15] Having completed secondary education corresponds to three years of schooling after compulsory school, which for the younger part of the sample lasted nine years. The measure takes into account the fact that the length of compulsory schooling has increased over time. Thus it may be regarded as a measure of an individual's level of education *relative to* his cohort.

[16] Before matching, 3829 observations were dropped because they stemmed from revisions. Observations totalling 108 concerned operations at foreign hospitals, for which waiting time is not registered.

patients' age and gender, and medical information specifically related to the hip replacement. Data on individuals' level of education, income, number of children and marital status are from the registers of *Statistics Norway*. These two registers can be perfectly merged by means of the unique personal identification code. *The Norwegian Patient Register* (hereafter NPR) has provided information on 46166 individual hospital stays within DRG 209, which includes hip replacements as well as other operations on hips, knees, ankles etc. Only the 25752 observations that had NSCP codes relevant for primary hip replacements were kept. For each hospital stay there are data on the patient's waiting time and home municipality, the name of the hospital, whether the stay was an emergency case or not, procedures executed, main diagnosis, secondary diagnosis etc. A *Matrix of distances* between all Norwegian municipalities provides information on driving distance by car in minutes, and makes it possible to identify the closest hospital given the patient's home municipality. It should be noted that travel distances within Norway are substantial in many cases. For long distances, flights are more relevant than the use of car, which we try to take into consideration in the model specification.[17]

Data from the NPR are merged with the NAR data using the variables patient's year of birth, gender, date of operation and hospital number. After matching, the combined data set consists of 19605 observations, which is 79% of the relevant component of the original NAR data set defined above.[18]

The following adjustments have been made: 682 observations were dropped because they are registered as emergency cases, for which the patient is not entitled to choose a hospital; 975 observations lacked information on when the patient entered the waiting list; 486 observations were dropped for fear of measurement error, as the reported waiting time was less than two days or more than 999 days;[19] 859 observations entered the waiting list on July 1 2003 or later and were dropped because data are truncated; and 6199 observations concerned patients who entered the waiting list before Free Choice of Hospital was introduced in 2001. Additionally, 592 observations were dropped so that each patient only

[17] Travel time by car is a more precise measure of distance than kilometers because the use of boats and ferries is taken into account when it is relevant. The distance is measured from the centre of one municipality to the centre of another.

[18] How well the two registers match varies among the institutions. Interest lies in whether some institututions are strongly under-represented or over-represented after the match compared to their share of operations in the NAR. Differences are traced, without any obvious explanation. The data set after matching is very similar to the before-matching NAR set with respect to mean and variation of sex, age and date of operation. One source of mismatch stems from the fact that bilateral hip replacements made during one hospital stay are counted as two observations with the NAR, but only one with the NPR.

[19] This exclusion criterion has been used in other studies of waiting times in Norway (The Office of the Auditor General of Norway (2003)).

has one observation in the sample.[20] Fifty observations were defined as leverage points. The criteria used for identifying leverage points are explained in the appendix. Finally, nine observations were dropped for other reasons.

1.5 Econometric framework

In principle, the allocation of operations could be thought of in a multinomial response setting where each of the 62 hospitals is regarded as a possible outcome, without any natural ordering. However, the question of interest is not which particular hospital is chosen, but rather the trade-off between specific hospital attributes, and whether attributes are valued differently depending on patient characteristics.[21]

To study how the trade-off $-\frac{dW}{dD}|_{dU=0}$ varies between patients on observable characteristics, we follow Tay (2003) and estimate a patient-level probabilistic choice model with interaction terms for patients' characteristics. The model to be estimated is:

$$U_{ih} = V_{ih} + \varepsilon_{ih}, \tag{3}$$

where ε_{ih} is an idiosyncratic patient–hospital error, which represents quality that is observable only to the patient and is treated as random. It corresponds to Z_{ih} in (2).

The functions f and g in (2) are assumed to be polynomials of degree m in D_{ih} and W_h, respectively. Therefore, V_{ih} can be specified as follows:

$$V_{ih} = \sum_{l=1}^{m}(\sum_{k=1}^{K}(\alpha_{0l} + \alpha_{kl}X_{ik})D_{ih}^{l} + \sum_{k=1}^{K}(\beta_{0l} + \beta_{kl}X_{ik})W_{h}^{l}) + \gamma q_h, \tag{4}$$

which is defined over all hospitals h = 1,.., H. Using a logit model, the probability that patient i chooses hospital h is given by:

[20] These cases concern patients who have two primary operations in the sample (one on each hip), which will be separate registrations in the NAR. They will appear as separate observations in the merged sample when the operations took place at different dates (during different hospital stays according to the NPR). Data for the oldest operation are retained.

[21] McFadden's choice model (McFadden, 1974), can be estimated by a conditional logit model, which in some respects is similar to a multinomial logistic regression. The models are suitable for different problems and have different data requirements: multinomial logit is intended for use when all that is known are the characteristics of the alternative chosen (and possibly the characteristics of the chooser), whereas conditional logit is suitable when we know the characteristics of the alternatives not chosen, as well. If all independent variables are attributes of the chooser, then the conditional logit model is exactly the same as multinomial logit (Stata reference manual, 2003). Both models share the Independence from Irrelevant Alternatives assumption. See Wooldridge (2002), p 500.

$$P_{ih} = \frac{e^{V_{ih}}}{\sum_j e^{V_{ij}}}.$$

The coefficients to be estimated are α_{0l}, α_{kl}, β_{0l} and β_{kl}, as well as the term γq_h. X_{ik} is patient *i*'s value for the patient characteristic *k*.[22] Thus, the marginal utility of waiting and of travel time is allowed to differ according to the patients' gender, age, level of education or the year they were placed on the waiting list.

The model can be estimated as a logit if one assumes that ε_{ih} is extreme-value i.i.d. (McFadden, 1974). The key assumption is that the errors are independent, which means that the unobserved portion of utility for one alternative is unrelated to the unobserved portion of utility for another alternative. To have independent errors, or ε_{ih} representing "white noise", V_{ih} must be well specified (Train, 2003, p 39). The concern is that there might be unobserved quality that is correlated with the observed quality regressors, so that the estimated effects will be biased. For instance, demanders could perceive long waiting lists at a hospital as a signal of high quality of treatment. To take into account unobserved quality, we have estimated the term γq_h in (2) by means of a dummy for each hospital. The dummy does not interact with patient characteristics. Implicitly, the effect of unobserved hospital-specific quality is assumed to be constant over the sample period, which is two and a half years. The time dimension enters the model through the element of the *k* vector that represents the year when the patient was placed on the waiting list.

In this case, it was not necessary to define a narrower choice set for computational reasons. The logit framework relies on the assumption that each ε_{ih} is independently, identically distributed extreme value. If this assumption is correct, the trade-off should be the same for two different choice sets.

If this assumption proves to be violated, so that the unobserved portion of utility is correlated over alternatives, a mix-logit model may be appropriate.

---

[22] For the reference individual, for whom X_{ik} =0 for all *k*, the expression simplifies to

$$V_{ih} = \sum_{l=1}^{m} (\alpha_{0l} D_{ih}^{l} + \beta_{0l} W_h^{l}) + \gamma q_h .$$

1.6 Results and discussion

To estimate (3) we tried different specifications of how waiting time and travel distance enter the model. For comparison, we estimated a very simple specification called "model A", where there are only linear terms for distance and waiting time. It seems to be a strong restriction to impose a linear relationship between utility on the one hand and waiting time or distance on the other. Therefore we shall focus on results from a quadratic specification ("model B") and a cubic one ("model C"). For ease of interpretation, only the interaction terms with the *level* form variables are included, i.e., α_{kl}, and β_{kl} in (4) are set equal to zero for $l > 1$. The signs of the estimated effects are the same in the different models, but the level of statistical significance varies somewhat, as can be seen from Table 2.[23] A non-linear relationship is especially motivated by the fact that when patients travel long distances, they will go by plane and not by car. Also, it cannot be ruled out that a patient will regard very short waits as inconvenient, because the long hospital stay and recovery period imply planning and making arrangements.

In the following, the overall effects of distance and waiting time on utility are discussed. Subsequently, separate effects of age, gender and education are commented upon, as well as how they change over time. Finally, we discuss the estimated trade-off for different patient categories (combinations of gender, age, education and year of referral) and compare our findings to other studies. The results are shown in Tables 2 and 3.

Distance and waiting time

Distance proves to have a significant negative effect on utility in both models. The variables concerning distance, which are distance, $(\text{distance})^2$ and, in model C, $(\text{distance})^3$, each turn out to be statistically significant at the 1% level, in estimations with a varying set of related interaction terms. The results with the full set of interaction terms are presented in Table 2. The disutility curve estimated in model C resembles the pattern of a cost curve; it is increasing for small travel distances, then flattens out or even falls, and becomes steeper and rising for high values.[24] With a quadratic utility function, the negative effect of distance is found to decrease for higher values of distance, so the disutility curve is concave.

---

[23] The interaction terms that are found to be statistically significant are the same in models A and B, except the interaction term between distance and gender, which is significant only in model A.

[24] The shape of the utility curve varies between patient categories. Here we shall only report the range where disutilty is increasing or decreasing for *all* cells. For the subsample with the lowest level of education, disutilty rises up to a travel time of 12 hours, then it decreases for distances between 16 and 19 hours and increases again for values of 23 hours or above. The subsample with more education is less reluctant to travel but shows a

Waiting time has an estimated negative effect on utility in both models. An F-test shows that the effect of the waiting times variables taken together is statistically significant at the 1% level both in models B and C. Similar to distance, the estimated marginal utility with respect to waiting time that model C yields is negative for all patient categories when estimated at mean values. Furthermore, the disutility from waiting rises over all values within the relevant range of waiting time. However, the economic significance of waiting time is small in terms of mobility (see Table 3).

Gender, age and education

Women are found to be less reluctant to wait and more reluctant to travel than men in both models, but the effect is not statistically significant, not even when the two gender interaction terms are tested together as a group. Old people are found to be less willing to travel than younger people, and the effect is statistically significant. The patient characteristic that proves to have the largest impact on preferences is level of education. Patients with more education are less willing to wait and more willing to travel, and the magnitude of the coefficients shows that the education effect on preferences is stronger than the age effect.

Change over time

There are several statistically significant changes from 2001 to 2003 showing more reluctance to wait and less reluctance to travel, but virtually no significant changes from 2001 to 2002. This result holds for both models. The change in the coefficients is relatively much larger for the waiting time variables than for the distance variables.

1.6.1 Discussion

Preferences for waiting and travelling clearly vary among patient categories. How willing they are to trade off a short distance for a shorter waiting time generally depends on where in the distribution of those variables the trade-off is measured. In Table 3, the trade-off in model C has been estimated at mean values for each patient category, after the sample has been split into two subsamples dependent upon patients' level of education. The estimated trade-off varies considerably between the two subsamples. At cell level, the MRS for the subsample with less education is 1.5 to two times higher than the sample with more education. For instance, the trade-off for the average individual in the reference group was 94 weeks in 2001,

---

similar pattern: disutilty increases up to a travel time of 11 hours, then it decreases for distances between 14 and 21 hours and increases again for values of 24 hours or above.

112 weeks in 2002 and 47 weeks in 2003, whereas the estimate for the same combination of age and gender but with more education was 52 weeks, 57 weeks and 32 weeks, respectively.

Within the cells belonging to the same year and subsample, there is a consistent ranking with respect to willingness to travel, as follows: younger men, older men, younger women and older women.[25] However, the most important factor in explaining the estimated MRS remains the level of education, as an older, more educated woman is more willing to travel than a younger, less educated man.

Within each subsample, the variation in the estimated trade-off is primarily along the time dimension. Table 2 shows that there is remarkably less reluctance to travel for shorter waits in 2003 than in 2001.[26] Note that the sharp decline in the MRS over time is partly because of the fact that the trade-off is measured at different values, particularly because mean waiting time has decreased for all patient categories in the period 2001–2003 (see Table 4). Within each year, the MRS is also estimated at somewhat different values for different cells. To isolate the time effect, we have estimated the MRS at the same values of waiting time and distance over the years, i.e., at mean values for each combination of gender, age and education. There is still a large difference between the MRS estimated in 2003 and in 2001 (not reported here).

However, the estimated change from 2001 to 2003 is not robust to changes in sample size. A robustness check using a somewhat smaller sample (9650 individuals instead of 9753, i.e., excluding those who travelled further than a distance within the first two quintiles of all possible distances) yields similar results concerning the effect of gender, education and age. However, the 2003 variables are not found to be statistically significant taken together in model C and the interaction term for the distance variable in 2003 has the opposite (negative) sign in both specifications. Thus, the result that there is a change in preferences over time seems to be driven by a small group of patients who chose extraordinarily long travel distances.

When we find that there are no statistically significant changes in revealed preferences over time for the large majority of the patient group, it could be because patients are not better informed of their rights in 2003 than in 2001 or because they truly prefer having the operation

[25] The exception is younger women in 2003.

[26] Concerning the estimate for 2002 it should be kept in mind that the change from 2001 to 2002 is not found to be statistically significant.

within a short travel distance. The two possible explanations cannot be disentangled, given the data we have.[27]

There is measurement error in the distance variable, as distance is measured at municipality level, from the (administrative) centre of the municipality where the patient lives to the centre of the municipality where the hospital is located. Thus, distances for within-municipality travel are set to 0, which is of course under-reported.[28] A more accurate measure, e.g., based on zip codes, is not available. The measurement error should be small if hospitals are located near the administrative centres and if patients have to go via their home municipality centre e.g., to reach major roads, train stations or airports. Then the distance variable could be interpreted as an extra travel distance, net of the distance to the home municipality centre. The measurement error is $e=x-x^*$, where x^* and x are the true and the measured value of an explanatory variable, respectively. Under the classical errors-in-variables (CEV) assumption that $Cov(x^*,e) = 0$, the magnitude of the estimated effect is underestimated (Wooldridge, 2003). In the case of within-municipality travel, the CEV assumption is violated because e and x^* are perfectly negatively correlated. Many patients chose a hospital within their home municipality, and for them the experienced distance is underestimated, which means that the reluctance towards distance appears to be greater than it actually is.[29]

We found that age and level of education influences preferences for this patient group. Other studies done within a national health system with waiting times show somewhat contradictory results, and different dependent and explanatory variables are included. Varkevisser (2006) finds that the likelihood of bypassing the nearest hospital decreases with age, while gender is insignificant. In a Danish study, Birk and Onsberg Henriksen (2005) found no statistically significant effects of age or gender on patient mobility[30], and a Norwegian enquete study in 2002 (Godager and Iversen, 2004) found that age is insignificant

[27] The studies by Godager and Iversen (2004) and Christensen and Hem (2004) shed light on how widespread information about patients' rights is, but they are single cross-section studies.

[28] About half of the patients in the sample live in municipalities that host a hospital and 32% of the patients chose to have the treatment within their own municipality.

[29] The across-municipality measurement error should be less than the within-municipality error. A separate estimation was made for a subsample of patients (4766 of 9753 patients) who did *not* have a hospital located in their home municipality. In both models B and C, the coefficients had the same signs and the same variables were significant as when we used the full sample.

[30] The sample is small (125 hip and knee patients responded to the questionnaire out of a study group of 144) and the maximum travel distance is 66 km, which is small within a Norwegian setting. Level of education was not included in the study.

and gender is significant only at the 10% level.[31] The latter study supports our finding that level of education matters. Gravelle *et al.* (2002) point out that education can be correlated with morbidity, the propensity to consult and the propensity to have private health insurance and use private hospitals. In our study, morbidity that is picked up by the age variable is corrected for at the individual patient level. The use of private health insurance and commercial hospitals for hip replacements is negligible in the study period.

Our finding that distance is a very important attribute for demand is supported for instance by Tay (2003) and, within a national health system framework, by Gravelle *et al.* (2002). The latter have examined the effect of waiting times on admissions, not hospital choice, and find a significant negative effect. Varkevisser (2006) analyses the decisions of two groups of Dutch patients to bypass the nearest hospital. He finds that extra travel time and low waiting time at the nearest hospital significantly decrease the probability of bypassing. The negative effect of extra travel time is much stronger for orthopaedic patients than for neurosurgical patients, who appear to put more weight on waiting time. Kjerstad and Kristiansen (2005) study 14 different DRGs and find large differences among the groups with respect to the probability of migrating given various covariates, among them age and gender; however, they did not control for waiting time differences among hospitals. Thus, patient group heterogeneity has to be taken into account.

1.7 Conclusion

Distance seems to be a very important attribute when patients consider hospital choice for elective hip replacements. Waiting time is also found to be statistically significant and to have a negative effect on utility, but the estimated effect, when it comes to behaviour, is found to be small. The fact that the marginal effect of waiting time on utility is estimated to be negative rules out the possibility that long waiting lists may be regarded as a signal of good quality. The model includes a hospital-specific fixed effect, which should cover time-constant effects like reputation.

The estimated trade-off varies considerably between models and patient categories. Patients are categorized according to age, gender, education and the year of referral. Avoiding distance is especially important to older people, and the estimates show no statistically significant gender differences. Clearly, the most important factor for the

[31] Their sample was patients with an unknown diagnosis who were on the waiting list or who had been hospitalized within the previous 12 months, who were asked to answer whether they "considered choosing a hospital themselves".

estimated marginal rate of substitution is level of education. Irrespective of age, gender and year of referral, a patient with more education is less reluctant to travel and less willing to wait. In the estimated sample, the mean patient of each category is less reluctant to travel for an operation in 2003 than in 2001, although this result is not robust to changes in sample size.

The most striking finding is the great reluctance to travel among patients having a primary hip replacement. The most mobility-inclined patient as measured by the marginal rate of substitution, a man under the age of 67 with higher education who entered the waiting list in 2003, must on average have a reduction in waiting time of 32 weeks to be willing to travel one extra hour.

When discussing the implications for health policy, caution must be exercised because the results refer to a specific patient group. Also, we cannot expect to see the full effect of the reform within the data period, which is two and a half years after its implementation. Given the data we have, we cannot decide whether low mobility is an expression of patients' preferences or is because of a lack of information on patients' rights and available alternatives. Still, the results indicate that the focus on waiting time in health policy might be overdimensioned.

References

Beregningsutvalget for spesialisthelsetjenesten (BUS) (2005) Årsrapport 2005, www.odin.dep.no (in Norwegian). [An expert committee appointed by the Department of Health, focusing on efficiency and effectiveness in the health care sector, annual report 2005.]

Birk, H. O. and L. Onsberg Henriksen (2003) "Hvilke faktorer bestemmer valget af sygehus?" *Ugeskr Læger* vol. 165(38), pp. 3608-3613 (in Danish). ["Which factors determine the choice of hospital?"]

Birk, Hans O. and L. Onsberg Henriksen (2006) "Why do not all hip- and knee patients facing long waiting times accept re-referral to hospitals with short waiting time? Questionnaire study" *Health Policy*, vol. 77(3), pp. 318-325.

Carlsen, B., T. Iversen and H. Lurås (2005) "Markedsforholdenes betydning for allmenlegenes praksisutøvelse", *Økonomisk forum* vol. 59 (4), pp.20-28 (in Norwegian). ["The impact of market conditions on the practise of general practitioners"]

Christensen, Ø. and K.G Hem (2004) "Fritt sykehusvalg i Norge", SINTEF Helse, report no. STF78 A015010 (in Norwegian). ["Free choice of hospital in Norway"]

Cullis, J., P.R. Jones and C. Propper (2000) "Waiting lists and medical treatment: analysis and policies", ch. 23 in Newhouse and Culyer (eds), *Handbook of Health Economics*, Elsevier.

Deaton, A. (2006) "Trying to be a good hip op consumer", *Royal Economic Society Newsletter*, Issue no. 133, pp. 3-4.

Department of Health, Norway, Revidert Nasjonalbudsjett 2005 (www.odin.no) (in Norwegian).

Directorate for Health and Social Affairs (2004) circular IS-12/2004 (in Norwegian).

Espehaug, B., O. Furnes, L.I. Havelin, L.B. Engesæter, S.E. Vollset and O. Kindseth (2006) "Registration completeness in the Norwegian Arthroplasty Register", *Acta Orthopaedica* vol. 77 (1), pp. 49–56.

Espehaug, B., L.I. Havelin, L.B. Engesæter and S.E. Vollset (1999) "The effect of hospital-type and volume on the survival of total hip replacements. A review of 39505 primary hip replacements reported to the Norwegian Arthroplasty Register, 1988–1996", *Acta Orthop Scand.* vol. 70(1), pp. 12–18.

Furnes, O., L.I. Havelin, B. Espehaug, L.B. Engesæter, S.A. Lie and S.E. Vollset (2003) "Det norske leddproteseregisteret - 15 nyttige år for pasientene og for helsevesenet", *Tidsskrift for den Norske Lægeforening* vol. 123(10),pp. 1367–1369 (in Norwegian). ["The Norwegian Arthroplasty Register – 15 years of service to the patients and the health care system"]

Godager, G. and T. Iversen (2004) "Hvem bruker retten til fritt sykehusvalg? Resultater fra Samordnet levekårsundersøkelse 2002", *Økonomisk forum* vol. 58 (4/5) (in Norwegian). ["Who make use of their right to free choice of hospital?"]

Gravelle, H., M. Dusheiko and M.Sutton (2002) "The demand for elective surgery in a public system: time and money prices in the UK National Health Service", *Journal of Health Economics* vol. 21(3), pp. 423-449.

Helse Nord (2003) "Organisering av elektiv ortopedi i Helse Nord, rapport fra arbeidsgruppe", www.helse-nord.no (in Norwegian). ["How elective orthopaedics is organized in the northernmost health region, report"]

Hoel, M. and E.M. Sæther (2003) "Public health care with waiting time: the role of supplementary private health care", *Journal of Health Economics* vol. 22(4), p 599–616.

Kjerstad, E. and F. Kristiansen (2005) "Free Choice, Waiting Time and Length of Stay in Norwegian Hospitals", SNF-Working Paper 21/05.

Losina, E., J. Barrett, N.N. Mahomed, J.A. Baron, and J.N. Katz (2004) "Early Failures of Total Hip Replacement. Effect of Surgeon Volume", *Arthritis & rheumatism*, Vol. 50(4), pp. 1338–1343.

McFadden, D. (1974) "Conditional logit analysis of qualitative choice behaviour", in P. Zarembka, ed. *Frontiers of Econometrics*, Academic Press, New York, pp. 105–142.

McGuire, T. (2000) "Physician Agency", ch. 9 in Newhouse and Culyer (eds), *Handbook of Health Economics*, Elsevier.

Nordsletten, L., L.I. Havelin, K. Indrekvam, A. Aamodt, S.E. Utvåg, K. Hviding (2002) "Valg av implantater ved innsetting av primær total hofteprotese i Norge", SMM rapport nr 6/2002, SINTEF Unimed (in Norwegian). ["Choice of implants for primary hip replacements in Norway", report]

Ot.prp. no 63 (2002–2003): *Om lov om endringer i lov av 2. juli 1999 nr 63 om pasientrettigheter m.m.* (in Norwegian).

Siciliani, L. and J. Hurst (2005) "Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries", *Health Policy* vol. 72(2), pp. 201–215.

Siciliani, L. (2005) "Does more choice reduce waiting times?" *Health Economics*, vol. 14(1), pp. 17–23.

Tay, A. (2003) "Assessing competition in hospital care markets: the importance of accounting for quality differentiation", *Rand Journal of Economics*, vol. 34(4), pp. 786–814.

The Norwegian Arthroplasty Register, annual report 2005, www.haukeland.no/nrl/

The Norwegian Knowledge Centre for the Health Services (2005) "Methodological development and evaluation of 30-day mortality as quality indicator for Norwegian hospitals", Report no. 4.

The Office of the Auditor General of Norway (2003) "Riksrevisjonens undersøkelse av effektivitet i sykehus, dokument nr. 3 (2002-2003)" (in Norwegian). ["Report on efficiency in hospitals"]

Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press.

Varkevisser, M (2006) "Why do patients bypass the nearest hospital? An empirical analysis for orthopaedics care and neurosurgery in the Netherlands", HEDG Working Paper 06/01, www.york.ac.uk

Vrangbæk, K., K. Østergren, U. Winblad Spångberg, H.O. Birk (2006) "Patients' reactions to free choice of hospital in Norway, Denmark and Sweden", HEB Working paper 01/06.

Wooldridge, J (2003), *Introductory Econometrics: A Modern Approach*, Thomson, South-Western

Wooldridge, J (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

www.curia.eu.int. (on the Yvonne Watts case).

Figures and tables

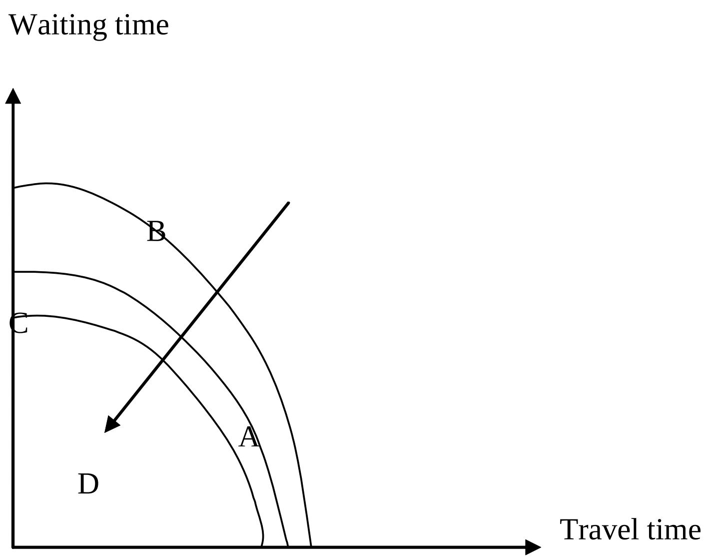


Figure 1. Patient preferences for elective surgery

Table 1. Descriptive statistics
for the alternatives actually chosen

Variable	Obs	Mean	Std. Dev.	Min	Max
Operation year	9753	2002.19	0.75	2001	2003
1 if placed on waiting list in 2002	9753	0.41	0.49	0	1
1 if placed on waiting list in 2003	9753	0.20	0.40	0	1
1 if female	9753	0.70	0.46	0	1
Age when placed on waiting list	9753	69.62	10.74	18	98
1 if age is above 66 years	9753	0.67	0.47	0	1
1 if have completed at least sec. education	9753	0.25	0.43	0	1
actual wait, days	9753	157.00	118.42	2	999
expected waiting time, weeks	9753	22.39	8.55	3	93.57
travel time, hours	9753	1.08	1.80	0	35.03
travel time to closest hospital, minutes	9753	29.25	45.92	0	465
1 if patient chose another hospital than the closest	9753	0.41	0.49	0	1

There are 596986 observations (combinations of hospitals and patients) and 9753 individuals in the sample. The table shows the values for the alternatives actually chosen.

Table 2. Hospital choice — estimated coefficients

	Model C			Model B		
Hospital choice	Coef.	Std. Err		Coef.	Std. Err	
Expected wait	–0.0374	0.0296		–0.0060	0.0100	
100*(Expected wait)^2	0.0498	0.0705		–0.0196	0.0099	**
1000*(Expected wait)^3	–0.0048	0.0048				
Distance	–2.3298	0.0356	***	–1.7544	0.0346	***
100*(Distance)^2	14.0524	0.3693	***	3.0366	0.0546	***
1000*(Distance)^3	–2.6641	0.1176	***			
Female interacted with:						
Expected wait	0.0033	0.0041		0.0025	0.0040	
Distance	–0.0214	0.0161		–0.0274	0.0212	
Old interacted with:						
Expected wait	–0.0002	0.0039		–0.0018	0.0038	
Distance	–0.0883	0.0169	***	–0.1627	0.0224	***
Year 2002 interacted with:						
Expected wait	0.0048	0.0062		0.0022	0.0054	
Distance	0.0189	0.0187		0.0658	0.0245	***
Year 2003 interacted with:						
Expected wait	–0.0166	0.0114		–0.0259	0.0082	***
Distance	0.0473	0.0207	**	0.0864	0.0248	***
Education interacted with:						
Expected wait	–0.0157	0.0047	***	–0.0162	0.0046	***
Distance	0.0985	0.0171	***	0.1025	0.0233	***
Number of obs		*596986*			*596986*	
LR chi2 (76)		*53579*			*52419*	
Prob > chi2		*0.0000*			*0.0000*	
Pseudo R2		*0.6676*			*0.6532*	

Single, double and triple asterisks indicate significant coefficients at the 10%, 5% and 1% levels, respectively. The models also include a dummy for each hospital (see equation 4).

Table 3. Hospital choice — estimated MRSs
for different patient categories, evaluated at mean values

	Less than secondary education				At least secondary education			
	Male		Female		Male		Female	
	<67 years	>67 years	<67 years	>67 years	<67 years	>67 years	<67 years	>67 years
MRS 2001	94	99	113	120	52	56	59	61
MRS 2002	112	116	141	151	57	61	65	69
MRS 2003	47	48	51	55	32	32	31	38

The marginal rate of substitution shows the reduction in waiting time, in weeks, needed to be willing to travel one extra hour. The estimates used are from model C and refer to equation (4). Mean values are reported in the table below.

Table 4. Mean values of distance and waiting times

	Less than secondary education				At least secondary education			
	Male		Female		Male		Female	
	<67 years	>67 years	<67 years	>67 years	<67 years	>67 years	<67 years	>67 years
2001, n =	*271*	*477*	*593*	*1605*	*205*	*188*	*247*	*280*
Distance	1.26	1.09	1.21	0.99	1.16	0.84	0.90	0.80
expected wait	26.19	26.43	26.14	26.06	25.30	24.66	23.96	23.41
2002, n =	*251*	*498*	*578*	*1643*	*196*	*242*	*246*	*316*
Distance	1.41	1.38	1.31	0.99	1.14	0.74	0.99	0.73
expected wait	23.21	23.03	22.91	22.86	21.12	20.29	21.72	20.97
2003, n =	*109*	*231*	*264*	*803*	*96*	*119*	*120*	*175*
Distance	1.07	1.36	1.29	0.98	1.36	1.36	1.85	0.78
expected wait	15.05	16.13	15.93	15.93	15.70	14.23	14.50	15.12

Expected wait is measured in weeks and distance is measured in hours of travel time.

Appendix

Leverage points

To detect leverage points, the choice set was divided into quintiles with respect to waiting time and distance, and the number of individuals who chose an alternative within each combination was counted. Observations that belonged to combinations with less than 10 individuals were dropped (50 observations). The table below shows the distribution of the 9803 observations that made up the data set before the 50 observations mentioned were excluded.

Appendix Table 1. Distribution of patients within the choice set

			Expected wait, quintiles					
			1	2	3	4	5	Sum
Distance, quintiles		*upper cut-off*	*15.8*	*19.5*	*24.5*	*29.2*	*93.6*	
	1	*3.3*	1950	2038	1893	1792	1421	9094
	2	*7.5*	134	119	81	172	50	556
	3	*9.8*	27	17	18	11	7	80
	4	*19*	18	7	7	9	5	46
	5	*53*	12	7	2	6	0	27
			2141	2188	2001	1990	1483	9803

Expected wait is measured in weeks and distance is measured in hours of travel time.

A short note on quality aspects of hip replacement

Total hip replacement is an operation designed to replace a hip joint that has been damaged most often by some form of arthritis, which causes pain, stiffness and deformity.[32] When arthritis has caused severe damage to the joint, a total hip replacement may be needed and the operation usually allows the patient to return to everyday activities (www.cdhb.govt.nz). This paper uses data on primary total hip replacements, which constituted 87% of all total hip replacements in 2004 (NAR, 2005).

A common procedure-specific measure of quality of total hip replacement is survival of the prosthesis, which refers to the duration from the primary operation until revision or until

[32] In 2004, 75% of the primary hip replacements were because of primary osteoarthritis. The second and third most common reasons for having the operation are fracture of the femoral neck and congentital dysplasia of the hip (The *Norwegian Arthroplasty Register*, 2005).

the patient dies or study closure. This paper focuses on primary hip replacements, but some patients may need a repeat operation of the hip replacement, most often because some of the components implanted have loosened.[33]

Total hip replacement is a type of surgery that is quite common. In Norway, more than 7000 operations take place every year (NAR, 2005) and most hospitals around the country can perform it. Still, there is some specialization among the hospitals. For instance, in Northern Norway, only two out of 11 hospitals offer revisions (Helse Nord, 2003), and complicated cases, where comorbidity often plays a part, are treated at university hospitals. The number of operations per year varies a lot among hospitals. The risk of revision is less in hospitals where surgeons perform a high number of operations per year (Espehaug *et al.*, 1999; Losina *et al.*, 2004). NAR has detected that some prostheses have a higher rate of revision. It has been recognized as a problem that surgeons use implants whose effect has not been documented clinically (Furnes *et al.*, 2003; Nordsletten *et al.*, 2002). Information on prosthesis survival related to individual hospitals or surgeons is not published in Norway, unlike in Sweden.

This study uses data on elective treatment. It is not obvious that data on mortality from emergency orthopaedic treatment are relevant for assessing the overall quality of the orthopaedics department or the hospital. A recent study shows that for some hospitals, the probability of death within 30 days from hip fracture is 65% greater than the average. The authors emphasize that data are not easily comparable and that this finding cannot be used to rank hospitals (*The* Norwegian Knowledge Centre for the Health Services, 2005).

---

[33] Of all hip protheses implanted in 1987–1990, 81% were still intact 16 years after the operation (NAR 2005).

Chapter 2

Education and Fertility: Evidence from a Natural Experiment*

by

Karin Monstad
Department of Economics
Norwegian School of Economics
karin.monstad@nhh.no

Carol Propper
Department of Economics
Bristol University, CEPR and IZA

Kjell G. Salvanes
Department of Economics
Norwegian School of Economics, Statistics Norway, Center for the Economics of Education (CEP) and IZA
kjell.salvanes@nhh.no

***Acknowledgements**
We are grateful to seminar participants at the Norwegian School of Economics and Business Administration and the Health Economics Bergen Programme for valuable comments on an earlier version of this chapter.

Abstract

Fertility continues to be an issue of public concern, even in developed countries that have experienced demographic transition and reached a state where both mortality and birth rates are low. Often when low birth rates and fertility patterns are discussed, women's trade-off between childcare and education and employment opportunities are brought forward as one explanation. However, many factors influence decisions on fertility, education and employment, very likely including unobservable factors that cannot be controlled for. Thus, causation is difficult to establish. In this paper, we make use of an educational reform to trace the causal effect of education on fertility outcomes. Our data enables us to estimate the effect of education on the timing of births as well as completed fertility, including the probability of being childless, after allowing for cohort effects. As the cohorts we study were born between 1947 and 1958, our data includes the most recent generation of women with completed fertility histories. The results indicate that increasing education at the lower tail of the education distribution leads younger women to postpone first births, most remarkably away from teenage motherhood towards having the first birth in their twenties and, for a smaller group, even until an age of 35 to 40 years. This result cannot be explained as a mere "incarceration" effect, and we interpret it mainly as the result of increased human capital accumulation from the reform. However, while the length of education and various fertility outcomes are found to be highly *correlated*, the data do not support any strong *causal* relationship other than the postponement of first birth. In particular, evidence is not found that more education results in more women being childless or leads to women having fewer children.

2.1 Introduction

Fertility continues to be an issue of public concern, even in developed countries that have experienced the demographic transition and reached a state where both mortality and birth rates are low. Low population growth and higher dependency ratios are argued to strangle economic growth. Recent OECD projections suggest that, because of demographic changes, the growth rate of per capita income will decline from 1.7% to 1.1% by 2050 in European countries and from 1.7% to 1.2% in the United States (Turner *et al.*, 1998). Often when low birth rates and fertility patterns are discussed, women's trade-off between childcare and education and employment opportunities are brought forward as one explanation. The observed relationship between fertility and female education varies between different countries and time periods, but there is much empirical support for strong correlations (Schultz, 1997; Cochrane, 1979). However, many factors influence decisions on fertility, education and employment, very likely including unobservable factors that cannot be controlled for. Thus, causation is difficult to establish. [34] In this paper, we make use of an educational reform to trace the causal effect of education on fertility outcomes. [35]

Nordic countries have a relatively high fertility rate (Sleebos, 2003), but this is an imperfect measure of long-run fertility as it aggregates behaviour over cohorts and ignores the timing of births. With respect to population development that is sustainable, the major concern in Nordic countries is the increasing number of childless women and the fact that the younger cohorts of women are having fewer children (Skrede and Rønsen, 2006). Our data enables us to estimate the effect of education on the timing of births as well as completed fertility, including the probability of being childless, after allowing for cohort effects. As the cohorts studied were born between 1947 and 1958, our data includes the most recent generation of women with completed fertility histories.

We study the relationship between the education of women and three fertility outcomes: the timing of children; childlessness; and the number of children. Our data confirms the expected correlation between fertility outcomes and education: women with more education are more often childless; they have fewer children and postpone births. Despite these statistically significant correlations, we do not find evidence of a causal relationship between the length of education on one hand, and completed fertility or

[34] Educational policy is rarely implemented to change fertility, but may still have fertility consequences. The interesting question is whether there is an intrinsic conflict between having a more educated population and obtaining sustainable total fertility levels in developed countries.

[35] Black, Devereux and Salvanes (2006) use this reform to assess the impact on teenage motherhood, while this paper analyses the effect on total fertility and the timing of first births in general.

childlessness on the other, when using the reform as an instrument for education. Our main finding is that increased mandatory education lead to the postponement of births; there are fewer cases of teenage motherhood and more first births among women aged 35 to 40 years.

The paper unfolds as follows. In section 2, we outline the major elements of the support system for parents in Norway and provide a short overview of the literature on fertility decisions, particularly the relationship between fertility and education. Section 3 describes the change in compulsory schooling that is used as an instrument in this study. The identification strategy and the data are described in sections 4 and 5. The results are presented and discussed in section 6. Section 7 concludes.

2.2 Background information

2.2.1 Previous literature

Economists model fertility in terms of the costs and benefits of children. There is a well-established literature describing how different types of costs affect fertility. Education may affect fertility decisions through several channels; in some contexts through better knowledge of contraceptives or through educational activity being incompatible with pregnancy or taking care of small children. Of special interest is the effect that goes through the labour market. In theory, the predicted effect of a rise in female wages on completed fertility is unclear because it depends on the magnitude of the different substitution and income effects. However, the conventional prediction consistent with Becker (1960) and Willis (1973) is that as child-rearing is a time-intensive activity, higher wages that raise a woman's opportunity cost of time will lead her to want fewer children, but probably put more resources into each child's upbringing. Hotz *et al.* (1997) provide a review of static models on completed fertility.

Theoretical models on the timing of births present a trade-off between the greater pleasure of early births and the lower costs of later births with a focus on the latter.[36] Much attention has been given to models that consider lifetime earnings, consumption smoothing and career planning rather than current incomes and wages. The literature points to woman's career costs as the most important explanation in favour of postponing births. In addition to the direct wage loss during labour force withdrawals, there is a loss in the returns to human

[36] Happel *et al*. (1984) assume that there is no pure time preference associated with the household's "effective" number of children.

capital in later periods due to depreciation.[37] For a review of dynamic fertility models, see Gustafsson (2001).

Sleebos (2003) provides a broad picture of the development in fertility in OECD countries from 1970 to 2000. Among the stylized facts is a general increase in the mean maternal age at first childbirth, with the exception of the USA and Korea. Countries differ remarkably in the degree to which women who have postponed childbirth then give birth at higher ages. This recuperation effect is stronger in Nordic countries, France and the United Kingdom than in Southern European and a number of other Continental European countries.

With respect to Norway, Lappegård and Rønsen (2005) have studied trends in the timing of first births for women born in the period between 1955 and 1969. They use longitudinal data up to 2001 and estimate a hazard model where education is treated as a time-varying covariate. Education is studied in several dimensions, including activity, length and field. They conclude that being a student clearly delays motherhood, but that the effect of length of education primarily works through the prolonged participation in the educational system. Field of study is found to have a separate impact, and is interpreted as mirroring different educational and career aspirations. However, this study does not correct for selection, i.e., the possibility that omitted, probably unobservable, factors influence both educational and fertility decisions.

To identify the causal effects of education, several studies have employed rules and regulations concerning school entry or dropout. An early and important contribution to the "natural experiment" literature is given by Angrist and Krueger in their 1991 paper where they used the quarter of birth as an instrument for educational attainment in earnings equations. The quarter of birth is correlated with the length of education because pupils were allowed to leave school by 16 years of age. To study the effects of education on fertility outcomes, several recent papers have used the same source of variation. McCrary and Royer (2006) use the school entry date as an instrument, and data from Texas and California. Their sample is selected because their source of data is birth certificates. They find no effect of mother's education on the timing of first births for women 23 years of age or younger.

Fort (2006) has utilized an Italian mandatory school reform enforced in 1963 that prescribed junior high school attendance, so that compulsory schooling increased from five to

[37] Gustafsson (2001) presents a list of parameters that will have a positive partial effect on the tendency to postpone births: the amount of pre-maternity human capital; the rate of depreciation of human capital due to non-use of human capital; the rate of return to human capital investment; and the length of time spent out of the labour force. In addition, the profile of human capital investments may play a role. Ignoring depreciation, theory has been ambiguous about whether a steep earnings function leads to earlier or postponed births (Cigno and Ermisch, 1989). Gustafsson argues that commonly used earnings profiles favour birth postponement.

eight years. The implementation period turned out to be unintentionally long and compliance was poor, especially in Southern Italy. This analysis does not, however, control for region; this raises/reintroduces the issue of selection.[38] As the author points out, economic conditions, traditions regarding fertility and labour market aspirations differed profoundly between regions. The estimated effects are restricted to those who had at most eight years of schooling. Her findings suggest that the reform lead these women to postpone their first childbirth, but they caught up with this delay in fertility before turning 26 years of age.[39]

Closest to our paper is work by Black, Devereux and Salvanes (2006) on the probability of teenage motherhood.[40] The authors use changes in compulsory schooling laws in the USA and Norway as an instrument for education in the two countries to identify the effect under two very different institutional environments. The Norwegian school reform will be described in detail later. Black *et al.* find evidence that increasing mandatory education reduces the incidence of teenage motherhood. Moreover, the size of the effect is estimated to be quite similar in the USA and Norway. Their results indicate that the effect of compulsory schooling laws goes beyond a pure "incarceration" effect. The current paper is an extension of Black *et al*. in that it studies fertility over the woman's fertile period, rather than just her teenage years, enabling us to examine the human capital effect of education on the full fertility history of women.

2.2.2 Institutional setting

Decisions on fertility are intertwined with decisions on marriage (or union) formation, education and employment. There have been changes over time in both norms towards single mothers, support systems for children such as day care and direct support, as well as the availability and acceptance of contraception. Skrede and Rønsen (2006) argue that what is regarded as Nordic "family policy" has not been aimed at fertility outcomes, e.g., sustainable total fertility levels, but rather at facilitating the combination of workforce participation and involvement in domestic tasks by both parents.

The support system for parents differs somewhat according to family type. For cohabitating or married mothers, the programs with the greatest implications for the cost of a child are the statutory universal rights parents have in connection with birth and the supply of

[38] Fort notes that sample sizes are too small to include region as a covariate. In addition, she lacks information on where the women lived at the time when they were around junior high school age.
[39] Effects of education on the timing of first childbirth are estimated for age levels 18 to 26 years, and are found to be statistically negative for ages 19 to 21 years.
[40] To our knowledge, this is the only existing paper that aims to estimate a causal relationship between fertility and education using Norwegian data.

subsidized childcare (discussed below) (Rønsen, 2004). A universal right to 12 weeks paid maternity leave was introduced in 1956, but the income compensation was relatively low. The major extension came in 1978 when maternity benefits were raised to cover 100% of the pre-birth income for 18 weeks.[41] This entitlement was extended to 20 weeks in 1987 and 22 weeks in 1988. Since 1993, mothers may choose between 42 weeks maternity leave with full pay and 52 weeks with 80% compensation. Since 1977, parents have been entitled to unpaid leave with job security until the child is one year old.[42] The cohorts in this study were in their peak fertility ages, i.e., 20 to 31 years old, when the first major extension came in 1978. However, the much more generous maternity leave reform in 1993 came too late to have any widespread impact as the women studied were then aged between 35 to 46 years.

The support system for single mothers is even more extensive, and from the early 1960s onwards, became very generous (Rønsen and Strøm, 1991). The system for single parents consists of several elements. The main part is a right-based support via the social security system ensuring single parents an income and temporary assistance to enable them to support themselves until the child is ten years old. This system was introduced in 1964 and became a part of the social security system in 1971. Together with other benefits, this enabled non-working single parents to take care of their children without working. The support is, by definition, temporary and is meant to support single parents (i.e., not living with the child's other parent, but may live with other partners). It is also meant to support single parents to become independent and provides support for education. All documented expenses for education are provided. Support was, and still is, income-dependent, i.e., reduced if the mother is working. The system was made less generous in 1998 (Skrede and Rønsen, 2006),[43] but by then "our cohorts" had virtually completed their fertility.

Another important element of the support system is income-dependent support for housing. Single mothers also receive financial support from the father if the father's name is registered with the authorities, and the authorities assist in enforcing child support payments. Single parents pay a reduced rate for day care (Rønsen, 2004). All parents receive a tax-free child allowance in Norway and single parents get about 1000 NOK extra per month in 2007. The attitude towards teenage mothers became more accepting in Norway during the 1970s

[41] To be eligible for maternity leave, the mother has to have worked for a certain period during pregnancy. From 1977, the requirement is six of the last ten months prior to birth. Alternatively, she gets a tax-free cash benefit at delivery, NOK 4730 in 1988 (Rønsen, 2004) and NOK 33,584 in 2007.
[42] Women working in the public sector can have longer unpaid leave, up to three years in total, but not less than one year per child. For instance, parents with three children are entitled to 3+1+1 years of unpaid leave.
[43] The duration of the support period was reduced from ten to three years in 1998. However, more incentives were given to work because the resulting reduction in the monthly support was not so drastic.

than before, which is also reflected in the fact that knowledge about sexual behaviour was made part of the compulsory school curriculum, and contraceptives (such as the pill) became more widely available. The pill was introduced in the late 1960s and was widely used; we know that among teenage girls aged 18 to 19 years in 1977, only 10% of those who had sexual intercourse did not use the pill or another type of contraceptive. In 1988, an even higher proportion used contraceptives (Noack and Østby, 1981 and Blom, Noack and Østbye, 1993). Abortion was legalized in Norway in 1979.[44] Although there is not one date that can be pinpointed when this started, the early to mid-1970s appears to be usually agreed upon (see Furre, 1992 for a general text on modern Norwegian history). These changes started within our period of analyses and again apply to the latter born, but not the earlier cohorts, within our data.

Public day care, which is subsidized in Norway, is subject to excess demand. Enrolment rates have risen sharply, from 5% in 1973 to 21% by 1980, 36% in 1990, 40% by 1992 and 54% in 2001 (Rønsen, 2004). The excess demand has been met by different forms of private childcare. A large proportion of Norwegian women work part time.

2.3 Compulsory schooling laws

In 1959, the Norwegian Parliament legislated mandatory school reform that increased the minimum level of education by extending the number of compulsory years of education from seven to nine years (thereby increasing the minimum dropout age from 14 to 16, as students start school at age seven). There were no exemptions to these laws. In addition, the reform standardized the curriculum and increased access to schools, as nine years of mandatory schooling was eventually made available in all municipalities.

The parliament mandated that all municipalities (the lowest level of local administration) implemented the reform by 1973. As a result, although it was started in 1960, implementation was not completed until 1972.[45] This suggests that for more than a decade Norwegian schools were divided into two separate systems; the system you were in depended on your year of birth and your municipality of residence. The first cohort that could have potentially been subject to the reform was that born in 1947. These individuals started school

[44] One would expect that access to legalized abortion may explain the drastic reduction in the number of teenage births in Norway from the late 1970s onwards, but in fact, the incidence of abortion has decreased, especially among teenagers from the early 1980s onwards (Lappegård, 2000).

[45] The reform had already started on a small and explorative basis in the late 1950s, but applied to a negligible number of students because only a few small municipalities, each with a small number of schools, were involved. See Lie (1974), Telhaug (1969) and Lindbekk (1992), for descriptions of the reform.

in 1954, and either finished pre-reform compulsory school in 1961, or went to primary school from 1954 to 1960, followed by post-reform middle school from 1960 to 1963. The last cohort who could have gone through the old system was born in 1958. This cohort started school in 1965 and finished compulsory schooling in 1972.

To receive funds from the government to implement the reform, municipalities needed to present a plan to a committee under the Ministry of Education. Once approved, the costs of teachers and buildings were provided by the national government. While the criteria determining selection by the committee are somewhat unclear, the committee wanted to ensure that implementation was representative across the country, conditional on an acceptable plan. (Telhaug, 1969, Mediås, 2000).[46] Figure 1 in the Appendix depicts the spread of the reform, focusing on the number of municipalities implementing the reform each year.

While it is not necessary for our estimation strategy, it would be useful if the timing of the implementation of the reform across municipalities were uncorrelated with teenage pregnancy rates, one of our outcomes of interest. To test this, we examine the relationship between the timing of the reform (by municipality) and teenage pregnancy rates prior to the reform (1960). We also look at other characteristics that may be associated with teenage pregnancy rates. For example, one could believe that poorer municipalities would be among the first to implement the reform given the substantial state subsidies, while wealthier municipalities would move at a much slower pace. However, work examining the determinants of the timing of implementation finds no relationship between municipality characteristics such as average earnings, taxable income and educational level, and the timing of implementation (Lie, 1973, 1974). Municipalities that are located geographically near municipalities that had already implemented the reform were themselves more likely to implement the reform; numerous interviews revealed that this was likely due to a particularly effective county administrator. As a result, the research supports a complex adoption process without finding support for a single important factor to explain the implementation process. To examine this issue ourselves, Figures 2, 3 and 4 in the Appendix depict the implementation of the reform by the average income, parental education and size of the municipalities. These figures suggest that there is little relationship between these factors and the timing of the implementation of the reform.

[46] Similar school reforms were undertaken in many other European countries in the same period, notably Sweden, the United Kingdom and, to some extent, France and Germany (Leschinsky and Mayer, 1990).

As a more rigorous test, in Table 1 in the Appendix, we regress the year of implementation on different background variables based on municipality averages, including parental income, the level of education, average age and the size of the municipality, as well as county dummies (there are twenty counties in Norway). Consistent with the existing literature, there appears to be no systematic relationship between the timing of implementation and parental average earnings, educational level, average age, urban/rural status, industry or labour force composition, municipality unemployment rates in 1960 and the share of individuals who were members of the Labour party (the most pro-reform and dominant political party).

2.4 Identification strategy

We study three fertility outcomes: the timing of children; the number of children; and childlessness. With one exception, the fertility outcomes Y_i studied are binary. Thus, the main specification used is a latent variable model:

(1) $$Y_i{*} = \alpha_0 + \alpha_1 R_i + \alpha_2 C_i + \alpha_3 M_i + e_i\,,$$

where Y_i = 1 if $Y_i{*} > 0$

Y_i = 0 otherwise.

The explanatory variables included are a reform indicator R_i, the set of municipalities *M* and cohorts *C,* which for individual *i* will take the value 1 for the municipality of residence and the person's cohort. The error term e_i is assumed to be i.i.d. and normally distributed $e_i \sim N(0,\sigma)$. A probit model is selected to estimate childlessness and the timing of births, while the number of children is estimated using ordinary least squares.[47]

There are a few points to note about eq. (1). To start with, it contains fixed cohort and municipality effects. The cohort effects are necessary to allow for secular changes in educational attainment over time that may be completely unrelated to compulsory schooling laws. The municipality effects allow for the fact that variation in the timing of the law changes across municipality may not have been exogenous to educational choice. Even if the reform was implemented first in areas with certain unobserved characteristics, consistent

[47] Regarding childlessness and the timing of births, OLS estimation results are reported for comparison purposes.

estimation is still achieved so long as: (a) these characteristics are fixed over time; (b) implementation of law changes are uncorrelated with changes in these characteristics; or (c) these characteristics are unrelated to the probability of the timing, the number of children, or childlessness.[48]

2.5 Data

2.5.1 Data sources

Based on different administrative registers and census data from Statistics Norway, a comprehensive data set has been compiled of the entire population in Norway, including information on family background, age, marital status, country of birth, educational history, neighbourhood information and employment information.[49] The initial database is linked administrative data that covers the entire population of Norwegians aged 0 to 90 years. This administrative data provides information about educational attainment, labour market status and a set of demographic variables (age, gender). To this, we match extracts from the censuses in 1960, 1970 and 1980.

To determine whether women were affected by the changed compulsory schooling legislation, we need to link each woman to the municipality where she grew up. We do this by matching the administrative data to the 1960 census. From the 1960 census, we know the municipality where the woman's mother lived in 1960.[50] The women we are using in the estimation are aged between two and 13 years in 1960. The indicator will be equal to one for a woman if, by her seventh year of schooling, the new system had been implemented in her municipality of residence, which is defined to be where her mother lived in 1960. One concern is that there may be a selective migration into or out of municipalities that implemented the reform early.[51] However, because the reform implementation did not occur before 1960, reform-induced mobility should not be a problem. A related concern is that random mobility at any point after we assign location may imply that an individual is not actually impacted by the reform, although we classify them as being so. This creates a

---

[48] Local variation in preference changes or shocks, e.g., regional differences in economic activity, is conceivable. As a robustness check, we estimated a model with municipality-specific trends.

[49] See Møen, Salvanes and Sørensen (2003) for a description of the data set.

[50] As very few children live with their father in cases where the parents are not living together, we should only have minimal misclassification through applying this rule.

[51] Evidence from Meghir and Palme (2005) for Sweden and Telhaug (1969) for Norway suggests that reform-induced migration is not a significant consideration.

measurement error problem that will tend to bias our estimates of the effects of the reform towards zero.

The measure of educational attainment is taken from a separate data source maintained by Statistics Norway. Educational attainment is reported by the educational establishment directly to Statistics Norway, thereby minimizing any measurement error due to misreporting. This register provides detailed information on educational attainment. The educational register started in 1970; for women who completed their education before then, we use information from the 1970 Census. Thus, the register data are used for all but the earliest cohorts of women who did not have any education after 1970. Census data are self-reported (four-digit codes of types of education were reported) and the information is considered to be very accurate; there are no spikes or changes in the education data from the early to the later cohorts.

Our primary data source on the timing of the reform in individual municipalities is from Ness (1971). To verify the dates provided by Ness, we examined the data to determine whether there appears to be a clear break in the fraction of students with less than nine years of education. In the rare instances when the data did not seem consistent with the timing stated in Ness, we checked individual municipalities by contacting local sources.[52] We are able to successfully calculate reform indicators for 672 of the 732 municipalities in existence in 1960. If the reform took more than one year to implement in a particular municipality, or we were unable to verify the information given in Ness (1971), we could not assign a reform indicator to that municipality. However, we have reform information for a large majority of individuals in the relevant cohorts.

We include those cohorts of women born between 1947 and 1958 in our sample. For these women, we observe their children in 2002. Thus, the youngest individuals will be 44 years of age and so all but a tiny minority will have completed their fertility. From the year and month of birth of the children and the year and month of birth of the mother, we can determine the age of the mother at birth to the nearest month. We exclude from our sample the small number of women who have a birth before they are aged 15 years and define a teenage birth as one occurring when the mother has not yet reached her 20th birthday. See table 2 in the Appendix.

[52] Between 1960 and 1970, a number of municipalities merged. In our analysis, we use the 1960 municipality as the unit of observation (Juvkam, 1999). In cases where the data were available at the 1970 municipality level, individual municipalities were contacted to determine the appropriate coding.

2.5.2 Descriptive statistics

Table 1 presents the key explanatory variables, the sample split by whether the individual were subject to the reform or not. The estimation sample consists of 290,604 women, 53% (154,818 individuals) of whom were affected by the reform. The subsample that lived in municipalities for which the reform was implemented had, on average, more education (a difference of 0.5 years). While the reform mandated nine years of schooling, the mean length of education for those *not* affected by the reform was 11.25 years; so many women had more than nine years of education even without the reform. The reform cohorts were born later (on average 4.5 years) as the reform was implemented gradually. The long-run trend to greater education explains at least part of the 0.5-year difference in average schooling.

The mean values in the lower part of the table show the distribution of cohorts within the reform and non-reform group, respectively. This reflects the gradual implementation of the reform; only 2.8% of those who were subject to the reform belonged to the three oldest cohorts (born between 1947 and 1949), while for the youngest cohorts the reform had been implemented for almost everybody.

Table 2 presents the outcome variables, split by whether the individuals were or were not subject to the reform. The differences in means are small, but the following patterns can be seen. Within the reform group, there are more women who are childless and the average number of children is slightly lower. The probability of teenage motherhood (first birth before age 20) is very similar for the two groups, but the group subject to the reforms were less likely to give birth in the first half of their twenties, and had a higher propensity to give birth after the age of 30 years.

These data are the result of cohort change as well as reform status. To separate cohort and reform status, Figures 1 to 7 present the outcomes by cohort, splitting the data into those subject to the reforms and those who were not. Figure 1 shows the time trend in the mean probability of a teen birth. This is higher in the non-reform group for almost all cohorts as well as showing a rise (for both groups) for those born from the early 1950s onwards. Figure 2 shows a slight decline in the probability of having a first child when aged between 20 and 25 as the cohorts get younger, which is matched by a slight increase for these same cohorts in having a first child between 25 and 30. However, the differences in these outcomes by reform status are small. Figures 4 and 5 show the trend towards first births being delayed until women are in their thirties. Later-born women were more likely to have first births later, and there is a greater tendency towards this in the reform group. For the youngest cohort (those born in 1958), the probability of not having a first child until 35 years or older is 0.032 for those subject to the reform; the corresponding figure for those who were not subject to the reform is 0.025, a percentage difference of about 30%. Figure 6 shows the evolution of the average number of children over the cohorts. Total fertility is quite stable over time and there is little difference between groups.[53] Figure 7 shows the trend in the probability of being childless. This increases for both reform and non-reform groups over time, but increases more for the reform group than their untreated counterparts.

Overall, the raw data suggest that those subject to the reforms were more likely to delay first birth, resulting in a drop in teenage motherhood and an increase in first births for those women in their thirties. There is also possibly some indication that the reforms resulted in a higher probability of being childless, at least for the younger cohorts included in the treatment group.

2.6 Results

2.6.1 The effects of the reform on fertility

Table 3 presents our key results. Each coefficient is derived from a separate regression, each of which controls for municipality and year of birth. Row 1 examines the correlation between education and fertility. The estimates show the expected strong statistical relationship between the length of education and fertility. Women who have more years of schooling have a higher tendency to remain childless; they also have fewer children and the probability of a first childbirth among the age groups less than 25 years decreases with education. The correlations estimated are of a rather large magnitude. For example, the probability of remaining childless increases by half a percentage point for each additional year of education.

Our interest lies in the causal relationships from education, and so in the coefficients from models where the educational reform is used as an instrument for education. The second and third rows of the table present the probit and ordinary least squares (OLS) estimates respectively. These are, in fact, very similar. The reform mandating more years of schooling reduces the probability of a teen birth, and increases the chance of not having a first birth until after age 35. There is also a relatively large, though not statistically significant, increase in first births between 20 and 25 years of age. There is no statistical effect on completed family size. The final row repeats the analysis using two stage least squares (2SLS) results. Again, we find the impact of education is to delay first births into the twenties and late thirties, but it

[53] The 1958 cohort is an outlier. The number of non-reform women in that cohort is small (238 observations).

has no effect on completed family size. Therefore, the effect of the reform is essentially to delay child bearing. In contrast with the raw association with education, when controlling for possible endogeneity in the education variable there is no significant relationship between education and the number of children born to a woman or the probability of never having children.

Although the magnitude of the estimated effects on timing is small in absolute terms, some are considerable in relative terms. The decrease in the probability of teenage motherhood is 5% relative to the frequency of teenage motherhood in the whole sample, which is 0.166 ((0.008/0.166)*100% = 5%). At the population level, the estimated effect is that about 260 fewer women would become teenage mothers each year if all individuals were mandated nine years of schooling as opposed to seven years.[54] Likewise, the increase in probability of giving first birth aged 35 to 40 is nearly 8% ((0.002/0.026)*100% = 7.9%), which is equivalent to an increase in 70 more women having children at this age each year.

2.6.2 Robustness checks

It is possible that our results are not picking up the impact of the reform but of unobserved differences between municipalities over time. To test for this, the top four lines of Table 4 present the results allowing for municipality, year-of-birth indicators and the corresponding interaction terms to allow for separate municipality–year effects. Standard errors are adjusted for clustering at the municipality–year level. The results are robust to these additional controls. The OLS and probit estimates using the reform dummy are very similar to those in Table 3; the 2SLS estimates show a larger effect of education on number of children, but again this is not statistically significant.

The group most likely to be affected by the reforms are those who are aged 13 years. To further test that our results are driven by the impact of education and not some omitted time-varying change in tastes, we used a regression discontinuity approach and re-estimated eq. (1) on a sample restricted to girls who were aged thirteen within three (alternatively five) years before or after the year of reform implementation. These results are presented in the second and third block of Table 4 respectively. In these cases, the correlation between education and the various fertility outcomes in the OLS model are very similar to those in Table 3. The causal results in Table 3 are also supported. The signs of the estimated effects

[54] The proportion of first birth by age in the sample is as follows: 0.17 in age group 15–20 years, 0.39 in age group 20–25, 0.23 in age group 25–30, 0.08 in age group 30–35 and 0.03 in age group 35–40, while the remaining 11% of the sample are childless. The mean size of a cohort of women in the whole population is about 32,000 individuals for the years 1947–1958.

are as before,[55] although the standard errors are larger because of the smaller sample size. The reduction of the probability of teenage motherhood is statistically significant in the model with a +/– five-year span, but the increased likelihood of first birth at age 35 to 40 is not.

2.6.3 Discussion

We can think of two mechanisms through which education can affect fertility. First, schooling is an activity that may reduce the possibility of behaviour that may lead to pregnancy. This is often referred to as the "incarceration effect". Second, education is an investment in human capital and may affect both the timing of births and the number of children. The incarceration effect is, by nature, temporary. If opportunity costs influence fertility in a lifetime perspective, it must be through the human capital effect.

Black *et al.* (2006) argue that if there is an incarceration effect, the data should show an increase in first births at ages 16 and 17 years after the dropout age was raised from 14 to 16 years. They find the opposite. We use essentially the same data, and also find that there is no catch-up in first births in the 15 to 20 age groups. This result alone tells us that the reform lead to more than merely an incarceration effect.

Our main finding is that the reform resulted in a postponement of births away from very early births and towards first birth at a later age. Due to the unfavourable consequences of teenage births, the results from increasing education that we find should be regarded as positive. Furthermore, the data does not show any statistically significant effects of the reform on total fertility. The allegation that education inevitably leads to fewer children being born is not supported by our data. As a caveat, if more schooling makes women tend to postpone their first birth until the end of their fertile period, this may have unfavourable consequences in terms of the increased risk of problems with fecundity and the corresponding costs to individuals and, in a publicly funded system, to the health care system.

The effects measured are "local average treatment effects"; the reform only affects those who change their behaviour because of the reform, i.e., those who would have chosen seven or eight years of education if compulsory schooling had not been extended to nine years. It may seem far-fetched that these women should postpone their first birth until the age of 35 to 40 years; this is a statistically significant result. It is likely that there is a great deal of heterogeneity between women regarding how they respond to the reform in compulsory schooling. Our data indicate that, in most cases, where first births were postponed due to the

[55] An exception is the estimated impact of education on the number of children, where the effect is either positive or zero in Table 4 while it is estimated to be negative, but very small, in Table 3. However, in none of the cases is the effect found to be statistically significant.

reform, the first birth took place at age 20 to 25 or 25 to 30 instead, although the difference between the reform and the non-reform group is not statistically significant for these age groups.

Postponement of births may also be given a human capital explanation. [56] For women who place a priority on establishing themselves in the labour market as a full-time worker, it is less costly to postpone childbirth, provided the age-earnings profile is not too steep. It seems plausible that the reform could lead some women into a different "track" in life: having had more compulsory schooling, the women impacted by the reform may have invested more than they otherwise would in secondary education or on-the-job training. Their preferences regarding when to have children may have changed so that they want to postpone birth as long as possible for career reasons, but eventually, the biological clock sets a fecundity limit.[57]

2.7 Conclusion

Using an educational reform as an instrument for education, we are able to investigate the causal effect of education on fertility. The data indicates that increasing education at the lower tail of the education distribution leads young women to postpone first births, most remarkably away from teenage motherhood towards having the first birth in their twenties and, for a small but statistically significant group, until the age of 35 to 40 years. This result cannot be explained as a mere "incarceration effect", and we interpret it as mainly the result of increased human capital accumulation because of the reform. While the length of education and various fertility outcomes are found to be highly *correlated*, the data do not support any strong *causal* relationship other than the postponement of first birth. In particular, we find no evidence that more education results in more women being childless or leads to women having fewer children.

[56] Gustafsson (2001) summarizes the theory on timing of births and points to the main explanations as being career planning and consumption smoothing. She finds that the main parameters which have an impact on career costs are the amount of prematernity human capital, the rate of depreciation of human capital due to non-use, the rate of return to human capital investments, the profile of human capital investments and the length of time spent out of the labour force.

[57] In principle, we may think of institutional changes as shocks that may alter both timing and total fertility, for instance make childless women change their mind about having a child or not. Perhaps the 1993 extension of paid marental leave spurred fertility in the reform group among cohorts that were fertile, but yet childless then – and in our sample that would be the 1953-1958 cohorts, who would be aged precisely 35-40 years. To assess the effect of the 1993 reform, we would need total fertility data on younger cohorts that are not included in our sample.

References

Aakvik, A., K.G. Salvanes and K. Vaage (2003). "Measuring heterogeneity in the returns to education in Norway using educational reforms", CEPR Discussion Paper no. 4048. Under revision for *European Economic Review*.

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

Angrist, J.D. (2004). "Treatment effect heterogeneity in theory and practice", *The Economic Journal*, vol. 114 (March), pp. 52–83.

Becker, G. (1960). "An economic analysis of fertility", in: *Demographic and Economic Change in Developed Countries*, Universities-National Bureau of Economic Research Conference Series 11 (NBER), pp. 209–231.

Black, S., P. Devereux and K.G. Salvanes (2006). "Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births", NBER Working Paper No. w10911. Forthcoming in *The Economic Journal*.

Blom, S., T. Noack and L. Østby (1993). "Giftermål og barn - bedre sent enn aldri?", in *Sosiale og økonomiske studier* 81. Oslo-Kongsvinger: Statistisk Sentralbyrå.

Cigno, A. and J. Ermisch (1989). "A micro-economic analysis of the timing of births", *European Economic Review,* vol. 33, pp. 737–760.

Cochrane, S.H. (1979). *Education and Fertility: What Do We Really Know?* John Hopkins University Press, Baltimore.

Fort, M. (2006). *Education and the Timing of Births: Evidence from a Natural Experiment in Italy*, Sep. 15th version of ISER Working Paper no. 20-2005.

Furre, B. (1996). *Norsk historie 1905–1990: vårt hundreår*. Det norske samlaget [Norwegian History 1905–1990: Our Century].

Gustafsson, S., (2001). "Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe", *Journal of Population Economics,* vol. 14(2), pp. 225–247.

Happel, S.K., J.K. Hill and S.A. Low (1984). "An economic analysis of the timing of childbirth", *Population Studies,* vol. 38, pp. 299–311.

Hotz, V.J., J.A. Klerman and R.J. Willis (1997). "The economics of fertility in developed countries", in: Rosenzweig, M.R., Stark, O. (eds), *Handbook of Population and Family Economics*, vol. 1A. Elsevier Science, North-Holland, Amsterdam, pp. 276–348.

Juvkam, D. (1999). "Historisk oversikt over endringer i commune- og fylkesinndelingen", report no. 99/13, Statistics Norway.

Lappegård, T. (2000). "New fertility trends in Norway", *Demographic Research*, vol. 2(3), p.16.

Lappegård, T. and M. Rønsen (2005). "The multifaceted impact of education on entry into motherhood", *European Journal of Population*, vol. 21, pp. 31–49.

Leschinsky, A. and Mayer, K.A. (eds.) (1990). *The Comprehensive School Experiment Revisited: Evidence from Western Europe.* Frankfurt am Main.

Lie, S.S. (1973). *The Norwegian Comprehensive School Reform. Strategies for Implementation and Complying with Regulated Social Change. A Diffusion Study.* Part 1 and II. Washington, D.C., The American University.

Lie, S.S. (1974). "Regulated social change: A diffusion study of the Norwegian comprehensive school reform", *Acta Sociologica*, vol. 16(4), pp. 332–350.

Lindbekk, T. (1992). "School reforms in Norway and Sweden and the redistribution of educational attainment", *Journal of Educational Research*, vol. 37(2), pp. 129–49.

McCrary, J. and H. Royer (2006). *The Effect of Maternal Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth,* NBER Working Paper No. W12329.

Mediås, O. A. (2000). *Fra griffel til PC*. In Norwegian. (*From Pencil to PC*). Steinkjer: Steinkjer kommune.

Meghir, C. and M. Palme (2005). "Ability, parental background and education policy: Empirical evidence from a social experiment", *American Economic Review,* vol. 95 (1), pp. 414–424.

Møen, J., K.G. Salvanes and E.Ø. Sørensen (2003). *Documentation of the Linked Employer–Employee Data Base at the Norwegian School of Economics*. Mimeo, The Norwegian School of Economics and Business Administration.

Ness, E. (ed.). (1971). *Skolens Årbok 1971.* In Norwegian. (*The Primary School Yearbook 1971*.) Oslo: Johan Grundt Tanum Forlag.

Noack, T. and L. Østby (1981). "Fruktbarhet blant norske kvinner. Resultater fra fruktbarhetsundersøkelsen 1977". [*Fertility Among Norwegian Women. Results from the Fertility Survey 1977*.] Samfunnsøkonomiske studier, nr. 49. Oslo-Kongsvinger: Statistics Norway.

Oreopoulos, P. (2006). "Estimating average and local average treatment effects of education when compulsory schooling laws really matter", *American Economic Review*, vol. 96, (1), pp. 152–175.

Rønsen, M. (2004). "Fertility and public policies - Evidence from Norway and Finland", *Demographic Research*, vol. 10, pp. 143–170.

Rønsen, M. and S. Strøm (1991). "Enslige forsørgeres tilpasning mellom trygd og arbeid." [Single mothers' choice of work and welfare]. In: Hatland (ed.) *Trygd som fortjent? En antologi om trygd og velferdsstat.* Oslo: Ad Notam.

Schultz, P.T. (1997). "Demand for children in low income countries". In: M.R. Rosenzweig, O. Stark, (eds), *Handbook of Population and Family Economics*, vol. 1A. Elsevier Science, North-Holland, Amsterdam, pp. 349–420.

Sleebos, J. (2003). *Low Fertility Rates in the OECD Countries: Facts and Policy,* Working Paper No. 15, Directorate for Employment, Labour and Social Affairs, OECD Paris.

Skrede, K. and M. Rønsen (2006). "Hvor bærekraftig er den nordiske familiepolitiske modellen?" *Samfunnsspeilet,* vol. 19(2), pp.2-10 (in Norwegian only) [How sustainable is the Nordic family policy model?"].

Telhaug, A. O. (1969). *Den 9-årige skolen og differensieringsproblemet. En oversikt over den historiske utvikling og den aktuelle debatt.* In Norwegian. (The 9-years compulsory school and the tracking problem. An overview of the historical development and the current debate). Oslo: Lærerstudentenes Forlag.

Turner, D., C. Giorno, A. de Serres, A. Vourc'h and P. Richardson (1998). *The Macroeconomic implications of ageing in a global context,* Working Paper No. 63, Economic Department, OECD Paris.

Willis, R. (1973). "A new approach to the economic theory of fertility behavior", *Journal of Political Economy,* vol. 81, pp. S14–S64.

Figures and tables

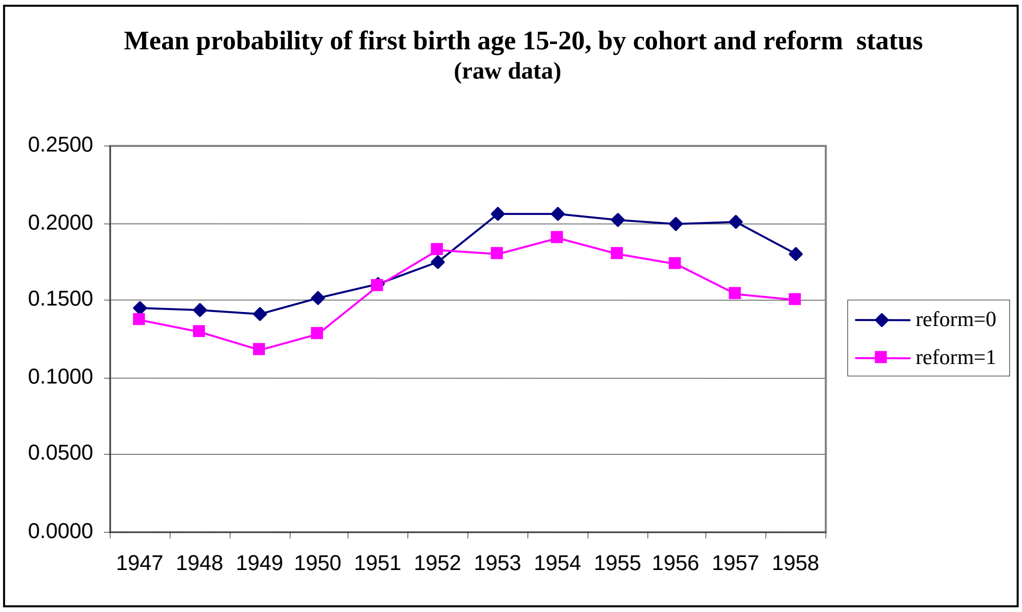


Figure 1

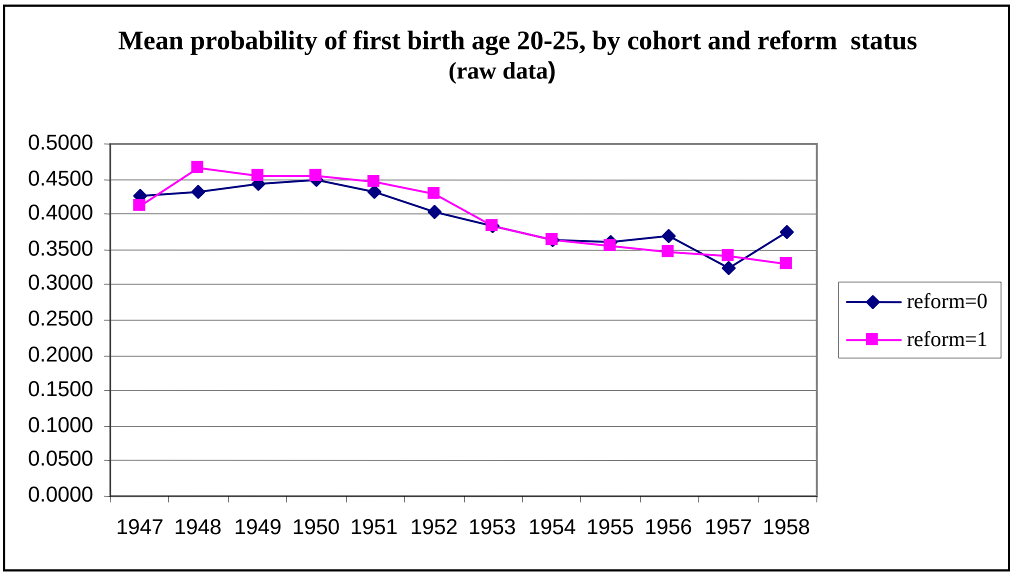


Figure 2

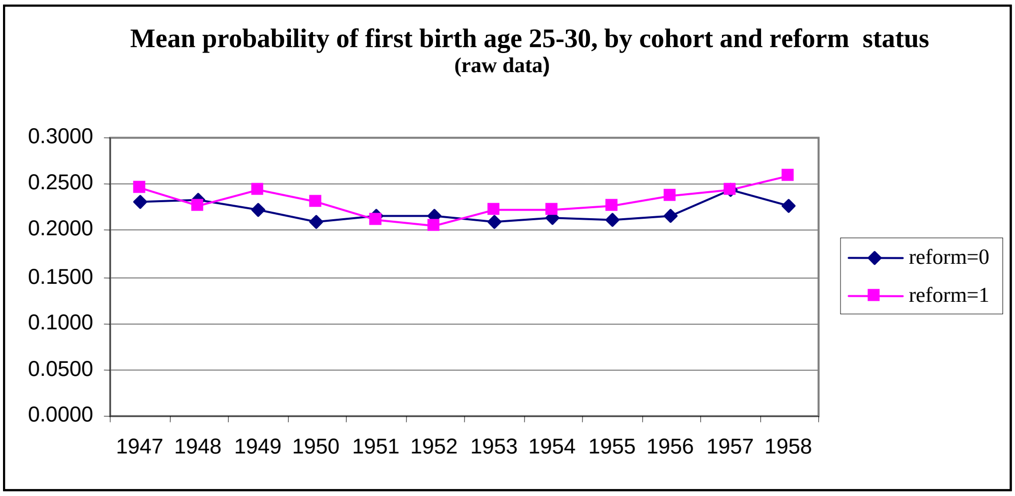


Figure 3

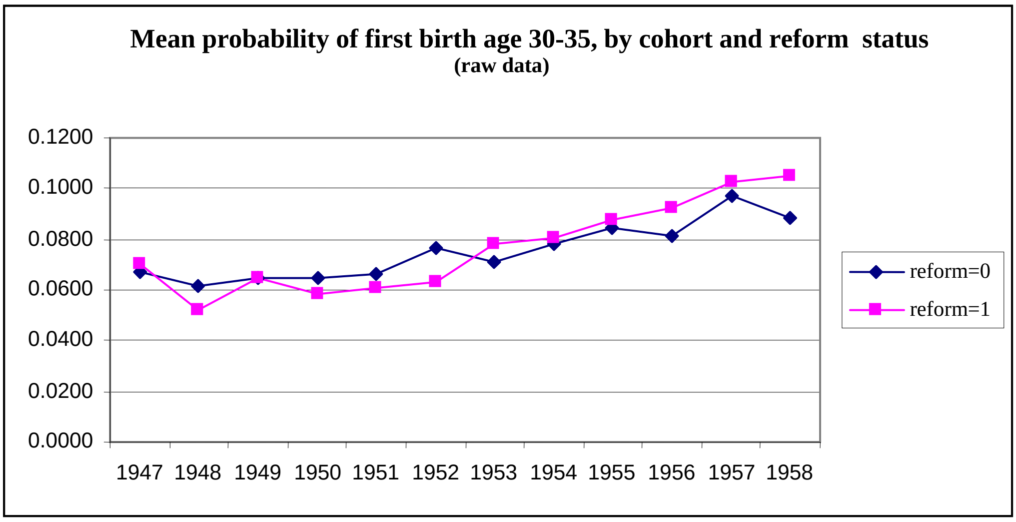


Figure 4

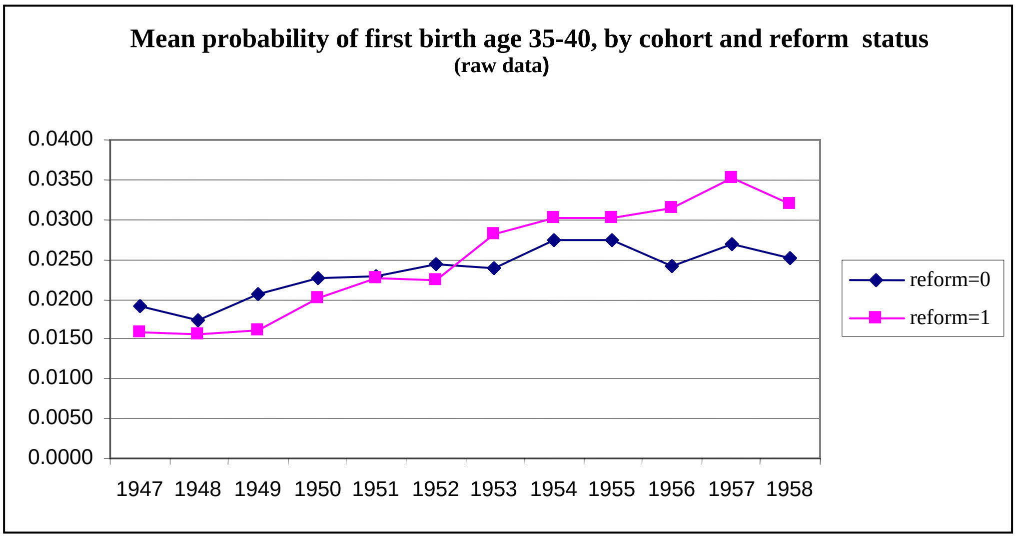


Figure 5

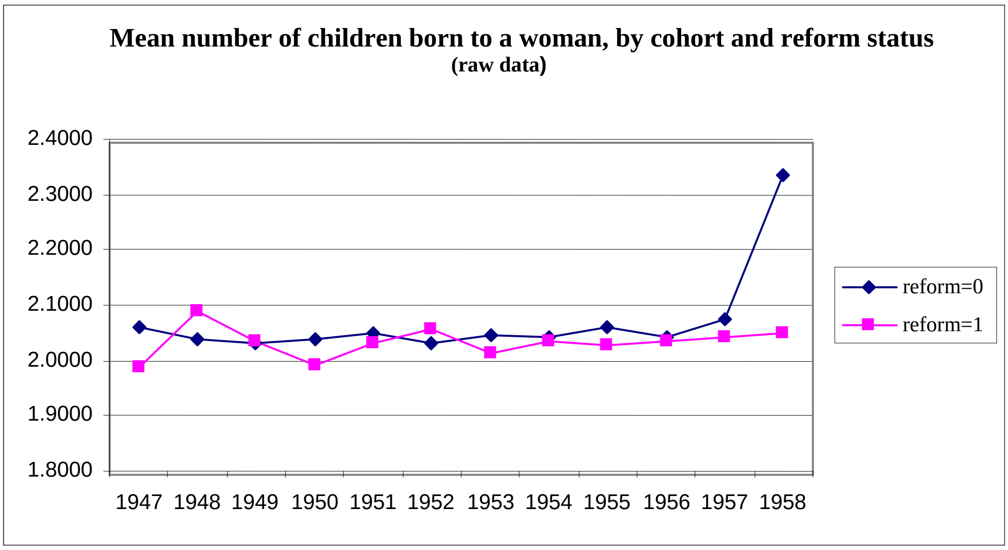


Figure 6

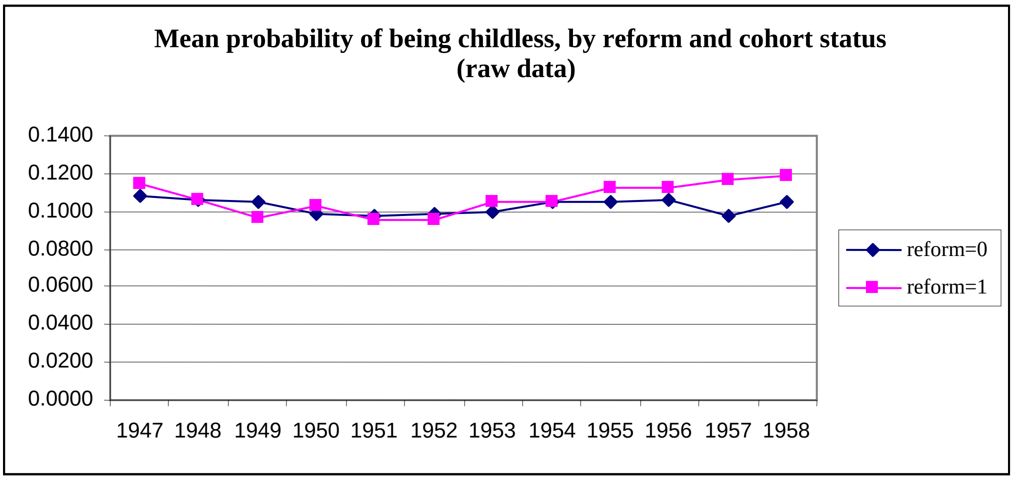


Figure 7

Table 1. Descriptive statistics, explanatory variables

	Reform=0					Reform=1				
	Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev	Min	Max
Years of education	135786	11.3	2.7	7	21	154818	11.7	2.5	7	21
Municipality	135786	1015.6	613.2	101	2030	154818	997.5	581.2	101	2030
Year of birth	135786	1950.6	2.6	1947	1958	154818	1955.1	2.4	1947	1958
Reform	135786	0.0	0.0	0	0	154818	1.0	0.0	1	1
Cohorts:										
1 if born 1958	135786	0.002	0.042	0	1	154818	0.176	0.381	0	1
1 if born 1957	135786	0.010	0.101	0	1	154818	0.169	0.374	0	1
1 if born 1956	135786	0.025	0.156	0	1	154818	0.159	0.365	0	1
1 if born 1955	135786	0.046	0.209	0	1	154818	0.137	0.344	0	1
1 if born 1954	135786	0.075	0.263	0	1	154818	0.110	0.312	0	1
1 if born 1953	135786	0.087	0.282	0	1	154818	0.099	0.298	0	1
1 if born 1952	135786	0.114	0.318	0	1	154818	0.064	0.245	0	1
1 if born 1951	135786	0.124	0.330	0	1	154818	0.039	0.193	0	1
1 if born 1950	135786	0.135	0.342	0	1	154818	0.022	0.146	0	1
1 if born 1949	135786	0.133	0.339	0	1	154818	0.012	0.107	0	1
1 if born 1948	135786	0.127	0.334	0	1	154818	0.009	0.092	0	1
1 if born 1947	135786	0.122	0.327	0	1	154818	0.007	0.081	0	1

Table 2. Descriptive statistics, outcome variables

	Reform=0					Reform=1				
	Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev	Min	Max
Outcome variables:										
1 if childless	135786	0.102	0.303	0	1	154818	0.110	0.313	0	1
Number of children	135786	2.044	1.095	0	14	154818	2.035	1.120	0	16
1 if first birth at age 15-20	135786	0.165	0.371	0	1	154818	0.167	0.373	0	1
1 if first birth at age 20-25	135786	0.417	0.493	0	1	154818	0.363	0.481	0	1
1 if first birth at age 25-30	135786	0.220	0.414	0	1	154818	0.235	0.424	0	1
1 if first birth at age 30-35	135786	0.069	0.254	0	1	154818	0.088	0.284	0	1
1 if first birth at age 35-40	135786	0.022	0.148	0	1	154818	0.030	0.170	0	1

Table 3. Results: Marginal effects of education

Expl. var.	Childless		Number of children		First birth age 15–20		First birth age 20–25		First birth age 25–30		First birth age 30–35		First birth age 35–40	
Schooling,	0.006	***	–0.013	***	–0.032	***	–0.024	***	0.030	***	0.015	***	0.005	***
OLS	(0.001)		(0.004)		(0.001)		(0.001)		(0.000)		(0.000)		(0.000)	
Reform,	0.001		–0.001		–0.009	**	0.005		0.001		–0.001		0.002	**
OLS	(0.002)		(0.010)		(0.005)		(0.004)		(0.003)		(0.002)		(0.001)	
Reform,	0.001		Irrelevant		–0.008	**	0.005		0.001		–0.001		0.002	**
Probit	(0.002)				(0.004)		(0.004)		(0.003)		(0.002)		(0.001)	
Schooling,	0.011		–0.009		–0.080	**	0.044		0.012		–0.008		0.021	**
2SLS	(0.018)		(0.087)		(0.039)		(0.032)		(0.028)		(0.018)		(0.009)	
N	*290596*		*290604*		*290604*		*290604*		*290604*		*290591*		*289057*	

Single, double and triple asterisks indicate significant coefficients at the 10%, 5% and 1% levels, respectively. Number of observations is reported for probit estimations, except for number of children, where n refers to an OLS model with reform as an explanatory variable.

The table shows the estimated coefficients from OLS estimations and marginal effects from probit estimations. Each column denotes separate regressions. Also included in the specifications are municipality and year-of-birth indicators. Standard errors are adjusted for clustering at the municipality level, confer eq. (1).

Table 4. Robustness checks

Expl.var.	Childless	Number of children	First birth age 15-20	First birth age 20-25	First birth age 25-30	First birth age 30-35	First birth age 35-40
(i)							
Schooling, OLS	0.006 ***	–0.013 ***	–0.032 ***	–0.024 ***	0.030 ***	0.015 ***	0.005 ***
Reform, OLS	–0.001	0.006	–0.008 ***	0.007 **	0.001	–0.002	0.002 **
Reform, Probit	–0.001	irrelevant	–0.007 **	0.006 *	0.001	–0.003	0.002 **
Schooling, 2SLS	–0.004	0.046	–0.070 ***	0.062 *	0.011	–0.021	0.022 **
(ii)							
Schooling, OLS	0.005 ***	–0.011 ***	–0.033 ***	–0.023 ***	0.030 ***	0.015 ***	0.005 ***
Reform, OLS	0.001	0.004	–0.006	0.006	0.002	–0.005 *	0.002
Reform, Probit	0.014	irrelevant	–0.005	0.006	0.002	–0.005 *	0.002
Schooling, 2SLS	0.001	0.041	–0.067	0.060	0.018	–0.051	0.022
(iii)							
Schooling, OLS	0.005 ***	–0.012 **	–0.033 ***	–0.024 ***	0.030 ***	0.015 ***	0.005 ***
Reform, OLS	0.001	0.000	–0.009 **	0.005	0.002	–0.002	0.002
Reform, Probit	0.001	irrelevant	–0.008 **	0.005	0.002	–0.002	0.002
Schooling, 2SLS	0.013	–0.004	–0.077 **	0.049	0.016	–0.016	0.016
n, (i)	*290596*	*290604*	*290604*	*290604*	*290604*	*290591*	*289057*
n, (ii)	*160044*	*160122*	*160075*	*160122*	*160122*	*159772*	*156875*
n, (iii)	*227188*	*227217*	*227217*	*227217*	*227217*	*226977*	*224970*

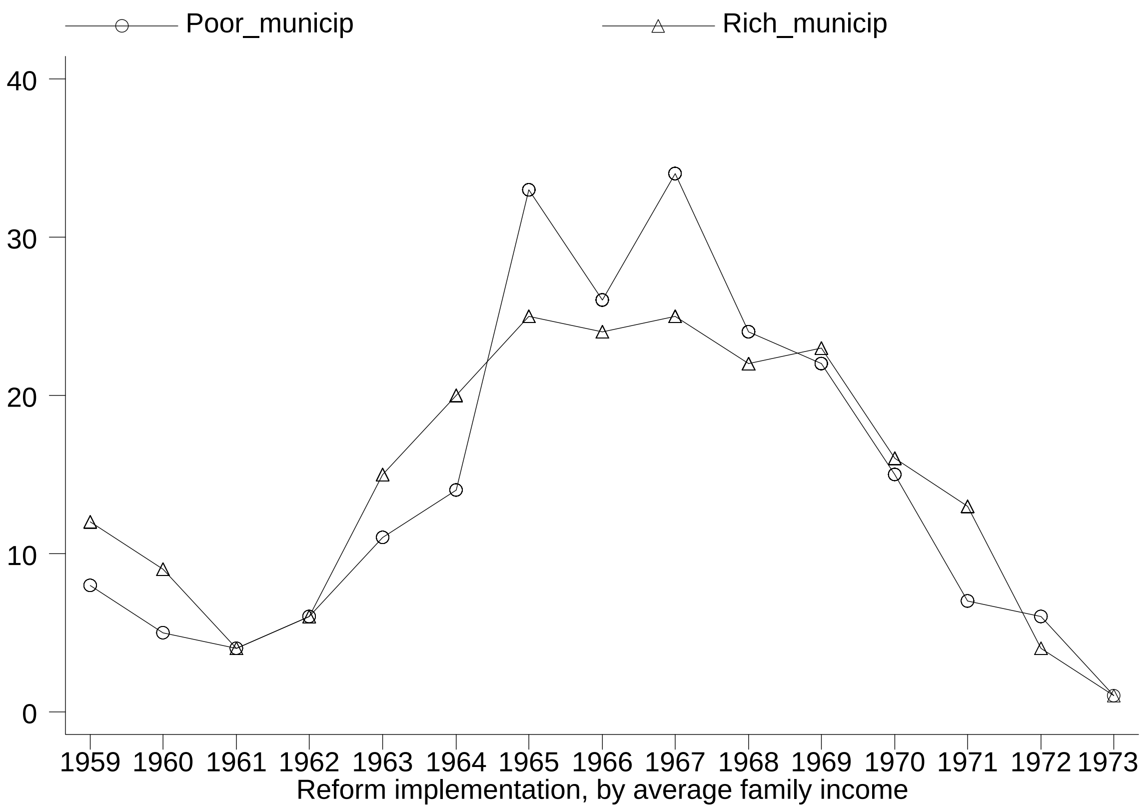
For general comments, see Table 3.

The table shows the results: (i) allowing for municipality-specific trends, (ii) Regression Discontinuity Approach with 3 year time span, (iii) with 5 year time span. The table shows the estimated coefficients from OLS estimations and marginal effects from probit estimations. Each column denotes separate regressions. Also included in the specifications are municipality and year-of-birth indicators. Standard errors are adjusted for clustering at the municipality-year level in estimation (i) and at the municipality level in estimations (ii) and (iii).

Appendix

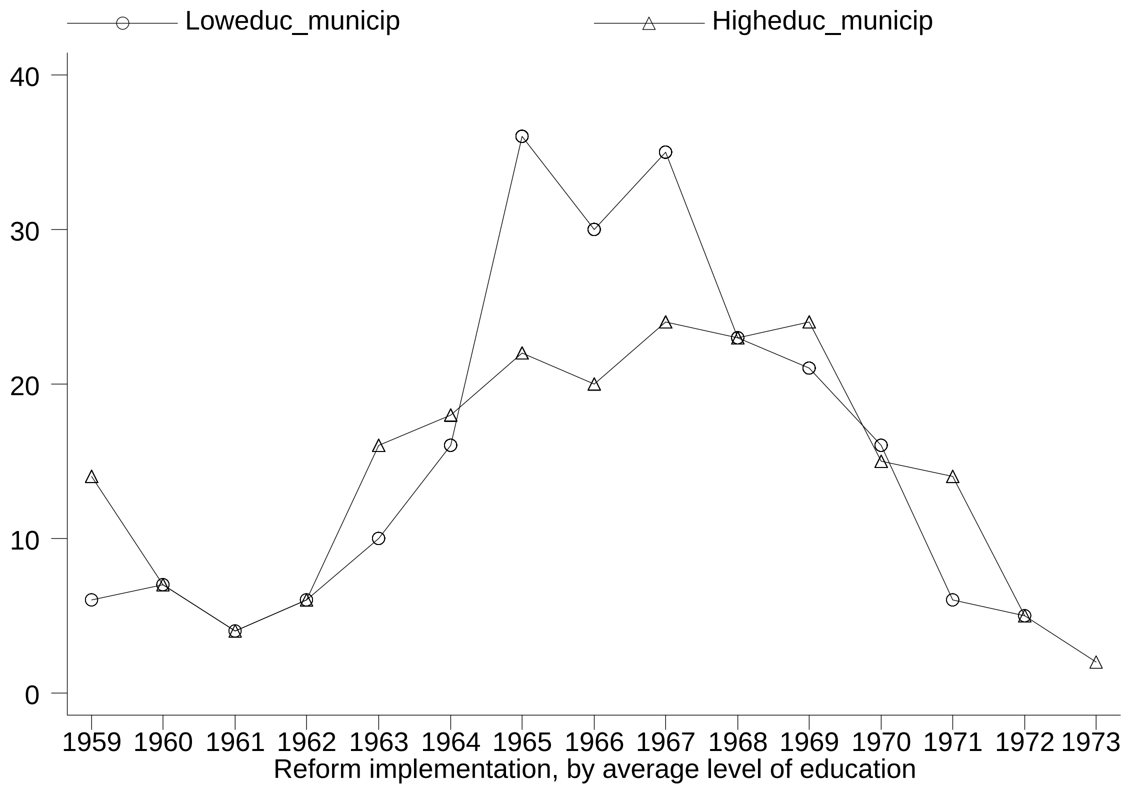
App. Figure 1
The Number of Municipalities Implementing the Education Reform, by Year

App. Figure 2
Reform implementation in Poor vs Rich Municipalities
Based on Average Family Income.



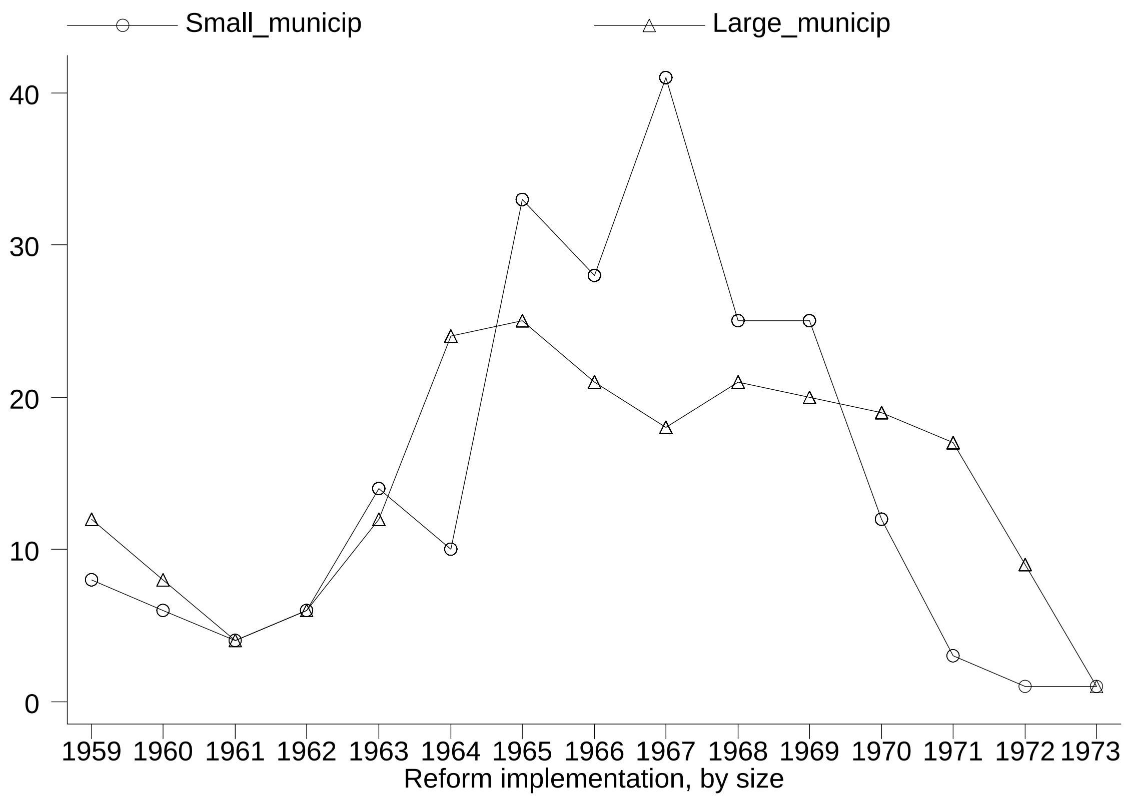
Poor (rich) municipality is calculated as below (above) median parent's income by municipality. Parent's average income is calculated for each municipality in 1970.

App. Figure 3
Reform Implementation in High vs Low Education Municipalities Based on Average Years of Father's Education in the Municipality



Low (high) education municipality is calculated as below (above) median education by municipality. Father's average years of education is calculated for each municipality in 1960.

App. Figure 4
Reform Implementation in Small vs Large Municipalities



Small (large) municipality is defined as below (above) median municipality as measured by population size in 1960.

Appendix Table 1

Timing of the Implementation of the Reform

Dependent Variable: Year of Reform	Coefficient	Standard error
County 2	–1.95	.65
County 3	5.02	5.23
County 4	–.64	.70
County 5	–.88	.67
County 6	–.90	.62
County 7	–1.21	.63
County 8	–1.90	.64
County 9	–1.21	.64
County 10	–2.20	.71
County 11	–.54	.63
County 12	–1.4	.60
County 13	–.45	.70
County 14	1.23	.59
County 15	–1.54	.58
County 16	.04	.60
County 17	–1.21	.57
County 18	–.26	.65
County 19	–2.77	.71
Share of Fathers with Some College	.92	3.88
Share of Mothers with Some College	12.30	8.31
Father’s Income (mean)	–.007	.004
Mother’s Income (mean)	–.01	.01
Father’s Age (mean)	.11	.16
Mother’s Age (mean)	–.12	.19
Size of Municipality/100	–.03	.03
Unemployment Rate 1960	–6.22	11.63
Share Workers in Manufacturing 1960	1.15	3.05
Share Workers in Private Services 1960	5.95	6.23
Share Labour Vote 1961	2.34	2.19
Constant term	1969.14	6.95

All variables are municipality level variables. Standard errors are adjusted for clustering at the municipality level.

Appendix Table 2
Data selection process

	Number of observations
Women born 1947–1958, in total	384385
Excluded because of motherhood before age 15	101
Excluded because woman's education <7 years	783
Missing on municipality	78952
Missing on reform indicator	11841
Missing on woman's length of education	2104
Sample size	290604

Chapter 3

Education and Fertility: Testing for Family Background and Spillover Effects*

by

Karin Monstad
Department of Economics
Norwegian School of Economics
karin.monstad@nhh.no

***Acknowledgements:**
I am grateful to Kjell G. Salvanes and Carol Propper for their valuable comments and discussion on this chapter. Thanks also to seminar participants at the Health Economics Bergen Programme and the Norwegian School of Economics and Business Administration.

Abstract

This paper analyses the effect of family background and social interaction on fertility choices over a woman's fertile period. The outcomes studied are the timing of first birth and whether women become mothers at all. I exploit a natural experiment—in the form of an educational reform—to correct for selection into education. The analysis benefits from a rich data set with information on parental education, age and income and the municipality of residence. In addition to examining parents' influence, I also investigate the impact of elder siblings of the same gender. Interest lies in how various aspects of family background interact with education, resulting in differences in fertility behaviour. Judging by the reaction to an increase in compulsory schooling, I find that the most important channel for the impact of family background on fertility is through family income and whether the young woman lives in a city. However, the potential spillover effect of the reform from elder to younger sisters is not found to be significant. The group that seems to have responded to the reform most strongly in terms of delaying first birth consists of women from low-income families, living in cities. The heterogeneity in responses is especially strong regarding the likelihood of first birth as a teenager. Thus, family background proves to be an important causal determinant for the effect of educational reform on fertility.

3.1 Introduction

In studies of fertility, it is a common finding that women's choice of education is an important explanatory factor (Kravdal, 1994; Hotz, Klerman and Willis, 1997). Studying the causal relationship between fertility and education, Monstad, Propper and Salvanes (2007) find that more education leads women to postpone first births, but that it does not result in lower total fertility or the greater incidence of childlessness. The causality is based on a natural experiment, i.e., an educational reform that increased compulsory schooling in Norway by two years. The effect estimated is by definition a "local average treatment effect" (Angrist, 2004); this of course raises the question about the generality of the results. Policy measures are often intended to benefit certain segments of the population, which is another reason to study heterogeneity in policy response. Indeed, one of the main aims of the educational reform in question, as stated explicitly in government documents, was to enhance the equality of opportunity along both socio-economic and geographic dimensions (Black, Devereux and Salvanes, 2005a). Furthermore, if education has a causal impact on fertility, particularly the timing of births, this is a potential channel through which education can have distributional consequences across generations.

Investment in education can be evaluated by the private rate of return. If externalities arise, the social and private rates of return will differ (Lucas, 1988). Even if educational reforms are hardly ever implemented because of their effect on fertility, one should bear in mind that such policy measures have fertility consequences and that fertility behaviour implies externalities. For instance, at the macro level, the number of children born and the age structure of the population have implications for economic growth. Research also suggests that teenage pregnancy shapes the life conditions for the child to be born in an adverse manner (for references, see Black *et al.,* 2006). Moreover, motherhood at a late age can have unfavourable medical consequences for the child: "…more stillbirths, more infant deaths, more premature births, more chromosomatic problems and more learning problems" (Gustafsson, 2001, p. 244).

One way that externalities may arise is that an individual's behaviour and norms may shape another person's preferences and behaviour. Such spillover effects are a special concern in the "new social economics literature" (Durlauf and Young, 2001). This literature examines such diverse phenomena as residential segregation (Schelling, 1971), neighbourhood effects on teenage childbearing (Crane, 1991) and how the presence of other smokers in a household affects the decision to quit smoking (Jones, 1994). Fertility is influenced by many factors, e.g., economic and cultural factors. It then appears reasonable that the family is an institution that shapes young girls' values and attitudes towards important decisions, including the choice of education and family formation. In several studies, the characteristics of the family have proven to have a great impact on young people's choice of education, labour market outcome, etc. (see e.g., Aakvik, Salvanes and Vaage, 2005; Black *et al*., 2005a and 2005b; Raaum, Salvanes and Sørensen, 2006). In this paper, I examine whether community and family background play an important role in decisions on fertility, and whether a spillover effect can be traced in the data. Elder relatives (grandparents, uncles and aunts) have been proven to have an impact on educational outcomes for same-gender adolescents (Loury, 2006). I will estimate the impact on fertility of elder sisters' education, while also controlling for the mother's and father's education.

When estimating social interaction effects, one of the challenges is to distinguish group influences (in this instance, sister influences) from any unobserved individual effects. I consider the possibility that growing up with a more educated sister reduces the propensity to become a teenage mother, conditional on other background characteristics, e.g., parental characteristics. The problem is that the sister's level of education is at least partially determined by parental characteristics, some of which are also unobservable. A natural experiment offers an approach to overcome this difficulty (Durlauf and Young, 2001).

The purpose of this paper is twofold. First, to examine the extent of heterogeneity in response to educational reform, and thereby identify the groups of women whose fertility behaviour changed the most owing to the reform. Second, to examine whether education triggers a spillover effect within the family, so that an elder sister's having more compulsory education has an impact on the younger sister's fertility outcomes, in particular the probability of teenage motherhood. Moffitt (2001) points to several methodological problems in identifying the effect of social interactions. This analysis benefits from a natural experiment; this helps solve the problem of unobservable heterogeneity. Unlike many other studies, the impact of family background is studied within the context where the link between education and fertility is causal.

The paper unfolds as follows. Section 3.2 gives a brief overview of the institutional setting and the compulsory schooling laws, as well as references to the relevant literature. The identification strategies chosen are presented in section 3.3 and the data sets used are described in section 3.4. The results are presented and discussed in section 3.5. Section 3.6 concludes.

3.2 Background information

In the literature on fertility choices, a woman chooses between two alternative uses of her time: participating in the labour market or taking care of children (Hotz *et al.*, 1997). Thus, studies on heterogeneity in the returns to education are relevant. Oreopoulos (2006) has addressed the question of heterogeneity from a broad perspective. Often it is claimed that educational reforms only affect the behaviour of a small part of the population, and that the results from studies using these reforms as instruments diverge from the average effect for the whole population. However, when Oreopoulos compares the effects of reforms of compulsory schooling across several countries, he finds that the estimated returns to education are very similar, whether they are estimated using reforms that affected almost half the population or only a small portion.[58] Using Norwegian data, Aakvik *et al.* (2005) have specifically studied the relationship between educational attainment and family background. The sample used is males and females born within the period from 1967 to 1972. The authors have data on family income at different periods of a child's life, which makes it possible to separate the long and short-term effects of income. They find that "…permanent income matters to a certain degree and that family income when the child is 0 to 6 years old is an important explanatory variable for educational attainment later in a child's life". The overall result is that "…long-term factors, such as permanent family income and parental education, are much more important for educational attainment than are short term credit constraints".

While Aakvik *et al.* (2005) study the impact of family background on education by means of a number of control variables, educational choice is still subject to selection because of unobserved factors. For instance, parental education can be positively correlated with parental ability and the ability of the offspring. I am able to examine the interaction between education and family background when there is an exogenous source of variation in education. Work by Oreopoulos indicates that the average and the local average treatment effects of education reforms are quite similar. Regardless, the mean effects may disguise substantial heterogeneity. To my knowledge, the observed heterogeneity in how women

[58] Oreopoulos (2006) focuses on the cross-country comparison of the mean effects. A number of socio-economic variables are used as control variables, but the differences in effects between socio-economic groups are not emphasized. The returns to schooling are estimated to be lower for males in most specifications. This finding holds across the countries studied, including the United States, Canada, the United Kingdom and Britain. Race is included in the model only for the US sample, and its impact depends on the specification employed.

respond to educational reform with respect to the timing of first births and childlessness has not been studied.[59] In this paper, this is analysed over a woman's entire fertile period.

The current analysis makes use of a compulsory schooling reform that the Norwegian Parliament legislated in 1959. This reform mandated that all Norwegians pupils attend two additional years of primary schooling (i.e., nine years) and was implemented by Norwegian municipalities at different times during the period from 1960 to 1972. For details on the reform itself and the implementation process, see Aakvik *et al.* (2003).

3.3 Identification strategies

The heterogeneity in the effect of the reform and the spillover effect are both identified by means of a difference-in-difference approach. Due to the structure of the data, the spillover effect is estimated using a subsample.

3.3.1 Identification strategy regarding heterogeneity analysis

Because interest lies in fertility outcomes Y_i that are binary, a probit model is used.[60] The main specification used is a latent variable model:

(1) $$Y_i{}^* = \beta_0 + X_i{}'\beta_1 + Z_i{}'\beta_2 + R_i Z_i{}'\beta_3 + e_i$$

where Y_i =1 if $Y_i{}^* > 0$ and Y_i = 0 otherwise, and where I define

$X_i{}' \equiv (R_i, C_i, M_i)$.

β_1 is a vector of coefficients for the set of individual characteristics X_i. The arguments of X_i are a reform indicator R_i, the set of municipalities C_i and cohorts M_i, which for individual *i* will take the value 1 for the municipality of residence and the reform person's cohort. Variation in the year of implementation among the municipalities makes it possible to control for both cohort and municipality when analysing the effects of the reform. β_2 is a vector of coefficients for the individual's background characteristics Z, where Z = (family income, mother's birth cohort, father's birth cohort, mother's level of education,

[59] McCrary and Royer (2006) include a control for maternal endowments in their analysis of the education effects on infant health. They comment: "…one could instead use an approximation that included interaction terms between schooling and endowments. Richer estimation equations such as these are, however, rare in the literature." Fort (2006) points to the problem of heterogeneity, but likely due to lack of data, does not examine how the effect of educational reform varies according to socio-economic characteristics.

[60] In the benchmark model, OLS estimation results are reported for the purpose of comparison.

father's level of education, urbanity). β_3 measures heterogeneity in the response to the reform, by means of the interaction terms $R_i Z_i$. The error term e_i is assumed to be i.i.d. and normally distributed, $e_i \sim N(0,\sigma)$.

In the benchmark model, all arguments in Z are set equal to zero, so the specification is

(2) $$\tilde{Y}_i^* = \alpha_0 + \alpha_1 R_i + \alpha_2 C_i + \alpha_3 M_i + \tilde{e}_i .$$

In this paper, I control for many aspects of observable heterogeneity. It should be noted that the reason to include background variables in eq. (1) lies in an interest in heterogeneity itself, and not to enable the better identification of the effects of R_i, as is sometimes attempted if there are concerns with endogeneity. The Norwegian mandatory schooling reform that I employ as an instrument for education in this paper, has been applied in other contexts by Aakvik *et al.* (2003), Black *et al.* (2005a, 2005b and 2006) and Monstad *et al.* (2007).[61]

3.3.2 Identification strategy regarding spillover analysis

The fertility outcome Y_i studied here is teenage motherhood of the younger sister in a group of sisters born within the reform cohorts. Thus, the main specification used is a latent variable model which is an extension of eq. (1):

(3) $$Y_i^* = \gamma_0 + X_i'\gamma_1 + Z_i'\gamma_2 + R_i Z_i'\gamma_3 + \varepsilon_i$$

where Y_i =1 if Y_i *>0 and Y_i =0 otherwise,

$X_i' \equiv (R_i, R_i R_i^S, C_i, M_i, D_i)$ and

$\gamma_1' \equiv (\delta_1, \delta_2, ...)$.

Equation (3) introduces the reform indicator R_i^S, which is related to the elder sister closest in age to the unit of observation *i*. R_i^S takes the value of 1 if the elder sister was

[61] In a natural experiment, the identification of the causal effect relies on the assumed source of exogenous variation being uncorrelated with any omitted variables that are correlated with the endogenous variable. The basic justification for the increase in compulsory education to be a natural experiment is the set up of the reform implementation. To demonstrate their point further, Black *et al.* (2006) regressed the year of reform implementation on a number of observable municipality characteristics and found no statistical significant relationships apart from the year dummies.

impacted by the reform, i.e., the mandated nine years of education. Thus, the variable of interest is $R_i R_i^S$. The model also includes as explanatory variables the set of municipalities M and cohorts C_i, and in most estimations the age difference between the sisters, D_i. The error term ε_i is assumed to be i.i.d. and normally distributed. Accordingly, a probit model is chosen for the estimation.

In principle, there are the following possible combinations of reform status for any pair of sisters:

- Case A: both the younger sister and the elder sister are impacted by the reform.
- Case B: the younger sister is impacted by the reform; the elder sister is not.
- Case C: neither the younger sister nor the elder sister is impacted by the reform.

The identification of δ_1 in eq. (3) utilizes variation in the younger sister's reform status, i.e., groups A and B compared to C. δ_2 is identified by means of variation between group A compared to groups B and C.

In the large majority of cases, there is only one sister for each individual in the sample, see Table 4 in the Appendix. When there is more than one possible pair of sisters, eq. (3) is estimated for the pair that is closest in age. The age difference D_i is defined accordingly.

3.4 Data

The analysis makes use of register data with information on all Norwegian women born from 1947 to 1958. To be included in the analysis, the woman's municipality of residence in 1960 and the reform status of the municipality must be known. The data set is very rich and includes background variables such as each parent's education, age and income. The income variable chosen is family income, defined as the sum of the mother's and father's income. For more information on the data set, see Monstad *et al.* (2007).

3.4.1 Data for heterogeneity analysis

After dropping observations because of missing information on background variables, the remaining data set consists of 274,581 observations. The data selection process is described in Table 1.

Within the restricted sample, 53% were affected by the reform. The descriptive statistics shown in Table 2a justify the argument that the effect of the reform must be considered a local average treatment effect. That is, while the reform mandated nine years of schooling, the mean length of education for those *not* affected by the reform was 11.26 years, so many women received more than nine years of education, even without the reform.

Regarding fertility outcomes, the non-reform group were subject to a pile-up of first births in the age group 20 to 25, while the age at first birth is more dispersed in the reform group.

The data show much variation in background variables, as can be expected given that a large part of the population is included. Differences in the year of birth and the years of education should be related to the fact that it took time to implement the reform: girls who were impacted by the reform are of a younger cohort than the non-reform group, and their parents are, on average, five years younger and better educated with 0.3 more years of schooling. The measure of parental education from the 1960 Census has been mapped onto the years of education following Raaum *et al.* (2006). Subsequently, parents are classified into three educational categories according to the length of schooling. There are many more men than women in the highest category defined, i.e., those with at least 12 years of schooling. In the reform group, a higher proportion lives in one of the ten major cities. Mean family income is considerably higher, which could be related to the higher level of education, wages being generally higher in cities and the presence of fewer old age pensioners among parents in the reform group. It should be kept in mind that within the parent generation, the level of education is generally quite low. More particularly, 55% of fathers and 65% of mothers in the sample have no more than compulsory schooling: that is, seven years of schooling. Only 9% of fathers and 2% of mothers received more than 12 years of education. There is a very strong correlation between the father’s income and family income (the correlation coefficient is 0.94), though the correlation between the father’s education and family income is much weaker (the correlation coefficient 0.41, see Table 1 in the Appendix).

Data on family income are taken from the 1970 Census. This is the data source closest in timing to the reform implementation. The impact of family income may change over a person's childhood and adolescence. Aakvik *et al.* (2005) have found that with regards to educational attainment, it is especially income in early childhood that matters. The income data in this study originate in one particular year, 1970, when the women in the sample were from 12 to 23 years old, with the mean individual aged 17 years. However, family income is strongly correlated over the life cycle, so I will use these income data as a proxy for income earlier in life.[62]

3.4.2 Subsample for spillover effect analysis

The data set consists of 48,574 observations of women who have at least one elder sister within the 1947 to 58 cohorts. Descriptive statistics are shown in Table 2b.

The population of sisters compares well with the larger population, see Table 2 in the Appendix. The most interesting aspect of the data set is the comparison between the three groups of women labelled A, B and C above. For each observation, the analysis uses two potential "treatments": first, being exposed to the reform yourself; and second, having an elder sister being exposed. The control group for the first treatment, group C, consists of women who were not impacted by the reform themselves, nor were their elder sisters. On average, these women are three years older and have less education, as expected. It took time to implement the reform, so the probability that two sisters have both been exposed to the reform is greater if they both belong to a younger cohort. For both to be in the 1947 to 1958 sample, with the younger sister belonging to group A, the age difference between them cannot be too large. Group B is defined in such a way that it includes many of the elder sisters from the older cohorts. Thus, the age difference between sisters within a family is, on average, 4.5 years in group B as compared to 2.8 years in group A. As a consequence, the sisters of group B members are, on average, 2.5 years older than group A's sisters.

It is noteworthy that the elder sisters of group A, on average born in 1953, had a much higher likelihood of teenage motherhood than the others (0.18 compared to 0.15 and 0.14). The data show a shifting trend in teenage motherhood. The frequency started to rise with the cohorts born in 1951 and 1952 and then fell from the 1955 to 1956 cohorts onwards: see Table 3 in the Appendix.

Equation (3) controls for both the younger sister's birth cohort and the elder sister's, through the age difference dummy.

[62] In principle, I could examine whether the impact of family income depends on the woman's age when income is measured. However, such a specification would introduce many more interaction terms and could become excessively complex.

3.5 Results and discussion

3.5.1 Results from heterogeneity analysis

As a benchmark, I estimated the effect of the reform without any interaction terms, and the results are reported in Table 3. The reform makes it less likely to have a first birth as a teenager and more likely to postpone birth until aged 20 years or above, with a statistically significant increase in the 35 to 40 years age group. The effect on childlessness is positive but statistically insignificant. These results are essentially the same as found when using a sample that is not restricted on background variables (Monstad *et al.*, 2007).

The results of including background variables are given in Tables 4 to 6. All three tables report the results from estimations of eq. (1), but family income is expressed by a whole set of quartile dummies in Table 4, and by a dummy for whether the family belongs to the bottom income quartile or not in Tables 5 and 6. Municipality dummies are also included, implying that fixed characteristics at the municipality level are controlled for, e.g., norms, average income level and local labour market conditions. The partial effects for these dummies are not reported. The additional background variables included are family income, the parents' year of birth and level of education and a dummy for whether the family lived in one of the ten major cities in 1960.[63] The base category is defined as follows: girls not impacted by the reform; those who come from low-income families where the parents are old[64] and belong to the lowest educational category; and who do not live in one of the major cities. Some of the background variables have strong *direct* effects on fertility, as can be read from the upper part of Table 4. However, the analysis will focus on the effect that goes via education, see the lower part of the table.

The overall picture when studying the response to the reform is that family income matters. Table 4 shows that the impact of family income is particularly strong for teenage motherhood. When compared to the bottom income quartile, the interaction terms for higher-level family income have positive signs, meaning that girls living in low-income families had

[63] Ideally, information on the parents' age at first birth would be useful. Unfortunately, such information is not available. As an alternative to using the parents' birth cohort, separate estimations for the mother's and the father's age when the child was born were undertaken. As this effect had the same sign but was of smaller magnitude, I chose to include the parents' birth cohort in the estimation.

[64] Fathers who belong to the oldest age quartile are born in 1914 or before and mothers in 1918 or before. The father's and mother's age when the child is born is on average 42.2 years and 37.6 years, respectively.

the highest reduction in probability.[65] Having a family income above the 1st quartile reduces the tendency of the reform to cause women to postpone first birth past the age range of 15 to 25 years, and it significantly weakens the response for childlessness. The variables representing father's education are dropped because of collinearity. This draws attention to the strong correlation between father's education, family income and mother's education shown in Table 1 in the Appendix. As shown, the family income categories most likely partially capture the effects of the father's education.

The mother's educational level proves to be an independent source of variation. Due to the reform, children of more educated and younger mothers and fathers tended to postpone first birth, not only past the teenage years, but also beyond ages 20 to 25. This impact is particularly strong if the mother has more than 12 years of education. Likewise, living in one of the ten major cities strengthens the effect of the reform in the direction of a decreased likelihood of giving birth as a teenager. Controlling for other variables, the reform also caused a small, but statistically significant, increase in the likelihood of being childless among urban women.

Family income and urbanity prove to be the most important background variables concerning the response to the reform, so I shall focus on these in the following discussion. Estimation with a full set of dummies for family income quartiles has shown that the effect for the bottom quartile is profoundly different from the other three quartiles. Therefore, I simplify the specification so that family income is expressed through a dummy indicating whether the family belonged to the bottom income quartile. Furthermore, the discussion will focus on the heterogeneity related to teenage motherhood. Teenage motherhood is the outcome variable on which the reform has proven to have the strongest estimated impact (see Table 3), and it is also the outcome where the heterogeneity in response to the reform is the greatest (see Table 4).

Table 5 shows the heterogeneity associated with income and urbanity over the whole fertile period, whereas Table 6 focuses on teenage motherhood and reports the heterogeneity with respect to income for urban and non-urban individuals separately. Table 5 further illustrates the finding that the reform had a greater impact on urban girls' tendency to give

[65] The magnitude of the positive partial effects for income quartiles 2, 3 and 4 may appear a puzzle because they are greater in size than the negative partial effect of the reform itself. Accordingly, it appears as if the net effect of the reform is positive for income quartiles above the lowest quartile. However, the magnitude of these partial effects is not comparable because they are computed at different values for the other variables (Wooldridge, 2003, p. 561). For instance, in computing the partial effect of the reform itself (–0.052), each income quartile is assumed to constitute approximately 25% of the population. In computing the partial effect of the interaction term with the second income quartile (0.146), it is assumed that income changes from the 1st income quartile as the base category to the second income quartile.

birth as a teenager. For the remaining outcome variables, the difference-in-effects between urban and non-urban girls are small. From Table 6 we can see that it is the poorer families within the urban community that respond most to the reform.

In most respects, the reform had an equalizing effect on the timing of births: the sign of the interaction term is the opposite of the sign of the background variable. This finding is generally true for family income and parents' age. Along the urban/non-urban dimension, the picture that Tables 4 to 6 provides is more mixed, because urbanity is linked with income.[66] Using a specification that focuses on the poorest income quartile, I find that the gap between urban and non-urban women is diminished because of the reform. On the other hand, the reform reinforced differences in fertility patterns according to the mother's level of education.

The finding that daughters of the most educated women respond so strongly to the reform is somewhat surprising, because one would think that girls from such families would be strongly encouraged to have an education at any rate, and that they would be less credit constrained than other groups. I interpret this result as an indication that the more educated mothers are, the more receptive they are to the general message of the reform: namely, that education is important for everybody. Through their own education or later career, these mothers may have become more oriented towards modern ideas. The reform is exogenous to marital ability, so if the daughters of well-educated women respond differently to the reform, it must be because of environmental factors, e.g., values and norms in their upbringing that correspond particularly well with the signal that the reform brings. Well-educated women are likely to advocate education for their daughters in general, and the educational reform seems to have helped stimulate their daughters further into postponing childbirth.

A clear result is that the reform had the greatest impact on women from low-income families. These individuals could be credit constrained or lack other resources at home, including stimulation, norms and role models that encouraged them to have an education beyond compulsory schooling or kept them from activities connected with a high risk of teenage motherhood. The estimated difference in the effect of the reform is quite dramatic:

[66] According to Tables 5 and 6, the difference between urban and non-urban women diminishes with the reform, whereas Table 4 provides the opposite picture. The result in Table 4 may be explained as follows: as poor women benefited most from the reform, but urban women are underrepresented within the lowest income quartile, the overall effect of the reform, as measured across all income quartiles, is to widen the gap between urban and non-urban women. Given that the main distinction in terms of fertility is between the lowest and the other income quartiles, the results in Tables 5 and 6 are far more interesting than those in Table 4.
In the estimation that Tables 5 and 6 are based upon, the effect of the urban variable itself is positive, whereas in Table 4 it is negative. This difference in signs stems from different ways of specifying the family income variable. It suggests that there may be different effects of being in the lowest income quartile (defined on a national basis) in a city than in a non-urban community. To avoid making the analysis too complex, I have not included interaction terms between urbanity and income.

while the probability of teenage motherhood is unchanged or slightly increased in the three upper income quartiles, it falls by 12 percentage points in the bottom income quartile, see Table 5. The change among the poorest is particularly strong in the larger cities (20 percentage points, as compared to 11 percentage points in rural municipalities or small towns, see Table 6). One possible explanation is that urban families who are poor compared to the national standard are relatively poorer than non-urban families, because the overall income level is higher in the major cities. Thus, poor urban families are negatively selected, and the reform has a stronger impact on young women's behaviour.

The reform lead both urban and non-urban women to postpone childbirth past the age of 25 years. This tendency cannot be interpreted as an "incarceration effect". According to human capital theory, it may be explained by the greater investment in women's education and the higher opportunity cost of her time (Monstad *et al.*, 2007). One possible reason why poor urban women react strongest to the reform could be that two additional years of compulsory schooling yields a higher return in a city because of the better labour market for women. Secondary and higher education is also generally more easily available in the cities. If the reform spurred some women into desiring further education, the lower cost of education in the cities could play a greater role after the reform than before.

3.5.2 Results of the spillover effect analysis

The direct effect of the reform, on the person exposed to it, is to decrease the likelihood of teenage motherhood, confer Table 3. The spillover effect measured by the interaction term δ_2 in eq. (3) must be interpreted as an additional effect of the reform, which may reinforce or weaken the negative effect.

The descriptive data indicate that age difference may be important in the analysis of spillover effects. One obvious reason is that the strength of a potential spillover effect could fade with the growing age difference; the closer in age sisters are, the more likely they are to share experiences, interests, friends, etc. Another reason is created by the natural experiment at hand, as the reform was implemented gradually. Trends in fertility behaviour also affect the elder sisters, and may have an impact on how they behave as role models. There are two similar ways of correcting for these trends: through an age difference variable as in eq. (3) or through indicators for the elder sister's cohort. In Table 7, four different models have been estimated.[67]

[67] All four estimations confirm the results previously displayed in Tables 3 and 4 that the reform reduces the probability of teenage motherhood for the mean individual.

For comparison, I have estimated eq. (3) without background variables (see the specifications labelled (I) and (II) in Table 7). The results for the variable of interest, the spillover effect, demonstrate that it can be important to control for age difference. In the model without an age difference variable, the spillover effect is positive and even statistically significant. The sign of the spillover effect turns negative once we control for age difference, which is what we should expect. That is, having an elder sister who has been mandated more education should set up a role model that makes younger sisters less inclined to become teenage mothers. However, the magnitude of the estimated effect is small, and the spillover effect is not statistically significant. In the specifications labelled (III) and (IV), I control for background variables as well. The main result is the same; the spillover effect is negative but statistically insignificant. A more complete picture of the estimation of eq. (3) is presented in Table 5 in the Appendix.[68]

3.6 Conclusion

In an earlier study, Monstad *et al.* (2007) found that a reform that enhanced mandatory education in Norway lead to the postponement of first births. In this paper, I examine to what extent it applies for different socio-economic groups, examining fertility over the whole of the women's fertile period. I also investigate whether an elder sister's reform status has any spillover effect on the younger sister's propensity to become a teenage mother.

Family background proves to be an important causal determinant for fertility behaviour in general, but also for the effect of educational reform on fertility. The analysis shows much heterogeneity in response to educational policy. In particular, the effect depends on family income and whether the young woman lives in a city. The heterogeneity in response is especially strong regarding the likelihood of first birth as a teenager. The group that responded to the reform most strongly in terms of delaying first birth consists of women from low-income families living in cities. These women also show an increase in the tendency to remain childless. However, the effect of family background does not seem to incorporate spillover effects of the reform from elder to younger sisters within the same family. The

[68] The table shows the partial effects of the background variables from eq. (3). These vary somewhat from the estimation without spillover effects, i.e., eq. (1). The decrease in teenage motherhood due to the reform is still greatest for women from low income families and those with young fathers. The interaction terms with urbanity and mother's education are no longer statistically significant. It should be kept in mind that the estimation is undertaken with a much smaller subsample, and that a relatively small proportion is classified as "urban" or have mothers with the highest level of education (13.8% and 2.4%, respectively).

spillover effect of the reform is estimated to have the expected sign (to reduce teenage motherhood), but it is small and statistically insignificant.

One of the main goals of the reform was to enhance the equality of opportunity along socio-economic and geographic dimensions. There was no objective stated with respect to differences in fertility patterns between socio-economic groups. Still, it is worth noting that as a consequence of the reform, the timing of first births and especially the frequency of teenage motherhood became more similar among the different income groups. Along the urban/non-urban dimension, the picture is more mixed. Using a specification that focuses on the poorest income quartile, I find that the gap between urban and non-urban women is diminished because of the reform.

References

Aakvik, A., K.G. Salvanes and K. Vaage (2003). "Measuring heterogeneity in the returns to education in Norway using educational reforms", CEPR Discussion Paper no. 4048. Under revision for *European Economic Review*.

Aakvik, A., K.G. Salvanes and K. Vaage (2005). "Educational attainment and family background" *German Economic Review,* vol. 6(3), pp. 377–394.

Angrist, J.D. (2004). "Treatment effect heterogeneity in theory and practise", *The Economic Journal*, vol. 114 (March), C52–C83.

Black, S., P. Devereux and K.G. Salvanes (2005a). "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital", *American Economic Review*, March, vol. 95(1), pp. 437–449.

Black, S., P. Devereux and K.G. Salvanes (2005b). "The more the merrier? The effect of family composition on children's outcomes", *Quarterly Journal of Economics,* May, vol. 120(2), pp. 669–700.

Black, S., P. Devereux and K.G. Salvanes (2006). "Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births", NBER Working Paper No. w10911. Forthcoming 2008 in *The Economic Journal*.

Crane, J. (1991). "The epidemic theory on ghettos and neighborhood effects on dropping out and teenage childbearing", *American Journal of Sociology,* vol. 96, pp. 1226–1259.

Durlauf, S.N. and H. Peyton Young (2001). "The New Social Economics", ch. 1 in *Social Dynamics,* S.N. Durlauf and H. Peyton Young (ed), Brookings Institution Press, Washington, D.C.

Fort, M. (2006). "Education and the Timing of Births: Evidence from a Natural Experiment in Italy", version of Sep 15th of ISER Working paper 2005-20).

Gustafsson, S., (2001). "Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe", *Journal of Population Economics* vol. 14(2), pp. 225-247.

Hotz, V.J., J.A. Klerman and R.J. Willis (1997). "The Economics of Fertility in Developed Countries" , ch. 7 in Rosenzweig, M.R., Stark, O. (eds), *Handbook of Population and Family Economics,* vol. 1A. Elsevier Science, North-Holland, Amsterdam, pp. 276-348.

Jones, A. (1994). "Health, addiction, social interaction and the decision to quit smoking", *Journal of Health Economics,* vol. 13, pp. 93–110.

Kravdal, Ø. (1994). "The importance of economic activity, economic potential and economic resources for the timing of first births in Norway", *Population Studies,* vol. 48, pp. 249–267.

Loury, L.D. (2006). "All in the extended family: effects of grandparents, aunts and uncles on educational attainment", *American Economic Review Papers and Proceedings,* vol. 96, pp. 275–278.

Lucas, R. (1988). "On the mechanics of economic development", *Journal of Monetary Economics*, vol. 22 (July), pp. 3–42.

McCrary, J. and H. Royer (2006). *The Effect of Maternal Education on Fertility and Infant Health: Evidence From School Entry Policies Using Exact Date of Birth,* NBER Working Paper No. W12329.

Moffitt, R. (2001). "Policy interventions, low-level equilibria, and social interactions", ch. 3 in *Social Dynamics,* S.N. Durlauf and H. Peyton Young (ed), Brookings Institution Press, Washington, D.C.

Monstad, K., C. Propper and K.G. Salvanes (2007). "Education and fertility: Evidence from a natural experiment", chapter in Doctoral thesis, the Norwegian School of Economics and Business Administration.

Oreopoulos, P. (2006). “Estimating average and local average treatment effects of education when compulsory schooling laws really matter”, *American Economic Review*, vol. 96(1), pp. 152–175.
Raaum, O., K.G. Salvanes and E.Ø. Sørensen (2006). “The neighbourhood is not what it used to be”, *The Economic Journal*, vol. 116 (January), 200–222.
Schelling, T. (1971). “Dynamic models of segregation”, *Journal of Mathematical Sociology*, vol. 1, pp. 143–186.
Wooldridge, J. (2003). *Introductory Economics: A Modern Approach*, Thomson South-Western.

Tables

Table 1. Data selection process

	Number of observations
Women born 1947–1958, in total	384385
Missing on cohort member's characteristics, or excluded:	
Excluded because motherhood before age 15 years	101
Excluded because woman's education is less than 7 years	783
Missing on municipality	78952
Missing on reform indicator	11841
Missing on woman's length of education	2104
	290604
Missing on background variables:	
Missing on father's education	7251
Missing on mother's education	239
Missing on mother's age	4029
Missing on father's age	2348
Missing on family income	2156
Sample size heterogeneity sample	274581
Subsample used in spillover effect analysis:	
Missing on mother's identification code	46433
The woman has no sister in the sample	136459
Dropped because is part of a group of triples	12
Sample of sisters	91677
The woman is the elder sister in the family, within the sample	43100
Dropped because less than 9 months interval between sisters' births	3
Sample of younger sisters used in spillover effect analysis	48574

Table 2a. Summary statistics, by reform indicator

	reform=0				reform=1			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Years of education	11.26	2.66	7	21	11.74	2.47	7	21
Municipality	1018.9	611.61	101	2030	997.7	580.23	101	2030
Reform	0	0	0	0	1	0	1	1
Year of birth	1950.7	2.57	1947	1958	1955.1	2.42	1947	1958
Background variables:								
1 if lived in one of the 10 major cities	0.15	0.35	0	1	0.22	0.42	0	1
Mother's education, years	7.94	1.64	7	18	8.19	1.78	7	18
1 if mother's education is 7 years	0.69	0.46	0	1	0.61	0.49	0	1
1 if 7<mother's education<= 12 years	0.29	0.45	0	1	0.36	0.48	0	1
1 if mother's education >12 years	0.02	0.14	0	1	0.03	0.16	0	1
Mother's age when daughter born	29.63	6.13	7	81	29.12	6.24	12	77
Mother's age in 1960	38.97	6.74	19	89	34.03	6.76	18	83
Father's education, years	8.65	2.53	7	18	8.97	2.65	7	18
1 if father's education is 7 years	0.58	0.49	0	1	0.51	0.50	0	1
1 if 7<father's education <= 12 years	0.33	0.47	0	1	0.38	0.49	0	1
1 if father's education >12 years	0.08	0.28	0	1	0.10	0.30	0	1
Father's age when daughter was born	33.22	7.03	0	86	32.63	7.02	1	87
Father's age in 1960	42.57	7.53	7	90	37.54	7.46	12	90
Family income in 1970, 100 NOK	260.43	286.58	0	14439	382.83	253.22	0	14058
Outcome variables:								
1 if childless	0.10	0.30	0	1	0.11	0.31	0	1
1 if first birth at age 15–20	0.16	0.37	0	1	0.17	0.37	0	1
1 if first birth at age 20–25	0.42	0.49	0	1	0.36	0.48	0	1
1 if first birth at age 25–30	0.22	0.41	0	1	0.24	0.42	0	1
1 if first birth at age 30–35	0.07	0.25	0	1	0.09	0.28	0	1
1 if first birth at age 35–40	0.02	0.15	0	1	0.03	0.17	0	1
N	*127733*				*146848*			

Table 2b. Summary statistics for subsample
Younger sisters used in the estimation of the spillover effect

Variable	Younger sister non-reform elder sister non-reform (Group C) Obs.	Mean	Std. Dev.	Younger sister reform elder sister non-reform (Group B) Obs.	Mean	Std. Dev.	Younger sister reform elder sister reform (Group A) Obs.	Mean	Std. Dev.
Expl. variables, younger sister:									
Years of education	12825	11.30	2.63	16721	11.70	2.48	19028	11.63	2.41
Municipality	12825	1118.41	597.51	16721	1100.69	603.88	19028	1048.04	567.58
Reform	12825	0.00	0.00	16721	1.00	0.00	19028	1.00	0.00
Year of birth	12825	1952.84	2.16	16721	1955.49	1.99	19028	1956.20	1.78
No. of sisters in family	12825	2.43	0.66	16721	2.27	0.54	19028	2.34	0.62
Age at first birth	11577	24.25	4.87	14967	24.55	5.03	16998	24.52	5.06
1 if first birth at age 15–20	12825	0.17	0.38	16721	0.17	0.38	19028	0.17	0.38
Information on elder sister:									
Age difference between sisters	12825	2.92	1.54	16721	4.49	2.16	19028	2.79	1.47
Year of birth	12825	1949.97	2.17	16721	1950.98	2.28	19028	1953.46	2.09
Age at first birth	11481	24.25	4.50	15033	24.24	4.61	17125	24.09	4.74
1 if first birth at age 15–20	12825	0.14	0.35	16721	0.15	0.36	19028	0.18	0.38

Table 3. Results benchmark model, without interaction terms

Explanatory variables	First birth age 15–20		First birth age 20–25		First birth age 25–30		First birth age 30–35		First birth age 35–40		Childless	
Length of education, OLS	–0.032	***	–0.024	***	0.030	***	0.015	***	0.005	***	0.006	***
	(0.001)		(0.001)		(0.000)		(0.000)		(0.000)		(0.001)	
Reform, OLS	–0.009	**	0.005		0.002		–0.001		0.002	**	0.001	
	(0.005)		(0.004)		(0.003)		(0.002)		(0.001)		(0.002)	
Reform, probit	–0.008	**	0.005		0.002		–0.002		0.002	**	0.001	
	(0.004)		(0.004)		(0.003)		(0.002)		(0.001)		(0.002)	
n in probit model	*274581*		*274581*		*274581*		*274570*		*272838*		*274574*	

Single, double, and triple asterisks indicate significant coefficients at the 10%, 5% and 1% levels, respectively**.** The table shows estimated coefficients from OLS estimations and marginal effects from probit estimations, confer eq. (2). Each column denotes separate regressions. Also included in the specifications are municipality and year-of-birth indicators. Standard errors are adjusted for clustering at the municipality level.

Table 4. Effects of reform, controlling for observed heterogeneity

	First birth aged 15-20 Partial effects		First birth aged 20-25 Partial effects		First birth aged 25-30 Partial effects		First birth aged 30-35 Partial effects		First birth aged 35-40 Partial effects		Being childless Partial effects	
Reform	–0.052	***	–0.040	***	0.040	***	0.013	***	0.007	***	0.033	***
	(0.008)		(0.010)		(0.008		(0.004)		(0.002)		(0.005)	
Background variables:												
Urban	–0.024	***	–0.011	***	0.024	***	0.018	***	–0.003	***	–0.013	***
	(0.003)		(0.003)		(0.002)		(0.001)		(0.001)		(0.002)	
Family income, 2nd quartile	–0.169	***	–0.044	***	0.169	***	0.059	***	0.017	***	0.086	***
	(0.003)		(0.009)		(0.006)		(0.004)		(0.002)		(0.004)	
Family income, 3rd quartile	–0.177	***	–0.042	***	0.195	***	0.057	***	0.019	***	0.072	***
	(0.004)		(0.010)		(0.006)		(0.004)		(0.002)		(0.006)	
Family income, 4th quartile	–0.188	***	–0.073	***	0.206	***	0.068	***	0.022	***	0.073	***
	(0.006)		(0.013)		(0.006)		(0.004)		(0.002)		(0.004)	
Mother's age, 1st quartile	0.065	***	0.042	***	–0.048	***	–0.023	***	–0.009	***	–0.030	***
	(0.005)		(0.006)		(0.005)		(0.003)		(0.002)		(0.004)	
Mother's age, 2nd quartile	0.033	***	0.023	***	–0.022	***	–0.009	***	–0.005	***	–0.017	***
	(0.004)		(0.004)		(0.004)		(0.003)		(0.001)		(0.003)	
Mother's age, 3rd quartile	0.017	***	0.016	***	–0.011	***	–0.003		–0.003	**	–0.014	***
	(0.003)		(0.004)		(0.003)		(0.002)		(0.001)		(0.002)	
Father's age, 1st quartile	0.068	***	0.031	***	–0.037	***	–0.019	***	–0.006	***	–0.031	***
	(0.006)		(0.007)		(0.005)		(0.003)		(0.002)		(0.004)	
Father's age, 2nd quartile	0.035	***	0.030	***	–0.022	***	–0.013	***	–0.004	***	–0.020	***
	(0.004)		(0.006)		(0.004)		(0.002)		(0.001)		(0.003)	
Father's age, 3rd quartile	0.025	***	0.014	***	–0.015	***	–0.008	***	–0.003	**	–0.012	***
	(0.003)		(0.004)		(0.003)		(0.002)		(0.001)		(0.002)	
Father's education,8-12 years	–0.046	***	–0.025	***	0.041	***	0.016	***	0.006	***	0.005	**
	(0.002)		(0.003)		(0.003)		(0.002)		(0.001)		(0.002)	
Mother's education> 12 years	–0.067	***	–0.099	***	0.066	***	0.032	***	0.008	***	0.015	**
	(0.008)		(0.010)		(0.012)		(0.005)		(0.004)		(0.006)	
Father's education, 8-12 years	–0.041	***	–0.019	***	0.042	***	0.016	***	0.006	***	0.005	***
	(0.002)		(0.002)		(0.002)		(0.001)		(0.001)		(0.002)	
Father's education > 12 years	–0.082	***	–0.091	***	0.075	***	0.042	***	0.014	***	0.021	***
	(0.002)		(0.005)		(0.005)		(0.003)		(0.001)		(0.004)	

(The table continues on the next page)

Table 4. Effects of reform, controlling for observed heterogeneity, cont.

	First birth aged 15-20 Partial effects	First birth aged 20-25 Partial effects	First birth aged 25-30 Partial effects	First birth aged 30-35 Partial effects	First birth aged 35-40 Partial effects	Being childless Partial effects
Interaction terms:						
Urban	–0.033 ***	–0.003	0.013 ***	0.006 *	0.000	0.014 ***
	(0.009)	(0.009)	(0.005)	(0.004)	(0.001)	(0.004)
Family income, 2nd quartile	0.146 ***	0.077 ***	–0.076 ***	–0.029 ***	–0.008 ***	–0.052 ***
	(0.016)	(0.013)	(0.008)	(0.003)	(0.002)	(0.004)
Family income, 3rd quartile	0.139 ***	0.075 ***	–0.080 ***	–0.023 ***	–0.008 ***	–0.046 ***
	(0.018)	(0.013)	(0.009)	(0.003)	(0.002)	(0.006)
Family income, 4th income	0.124 ***	0.093 ***	–0.067 ***	–0.022 ***	–0.007 ***	–0.048 ***
	(0.022)	(0.016)	(0.010)	(0.003)	(0.002)	(0.005)
Mother's age, 1st quartile	–0.019 ***	–0.008	0.033 ***	0.003	0.000	0.005
	(0.006)	(0.008)	(0.008)	(0.005)	(0.003)	(0.005)
Mother's age, 2nd quartile	–0.019 ***	–0.013 **	0.023 ***	0.001	0.001	0.004
	(0.005)	(0.007)	(0.006)	(0.004)	(0.002)	(0.005)
Mother's age, 3rd quartile	–0.011 ***	–0.005	0.010 *	0.000	–0.001	0.005
	(0.004)	(0.006)	(0.006)	(0.003)	(0.002)	(0.004)
Father's age, 1st quartile	–0.023 ***	–0.008	0.014 **	0.009 *	0.004	0.012 **
	(0.006)	(0.009)	(0.007)	(0.005)	(0.003)	(0.005)
Father's age, 2nd quartile	–0.017 ***	–0.019 **	0.015 **	0.009 **	0.002	0.008 *
	(0.006)	(0.008)	(0.006)	(0.004)	(0.002)	(0.005)
Father's age, 3rd quartile	–0.017 ***	–0.009	0.013 **	0.003	0.003 *	0.006 *
	(0.005)	(0.006)	(0.006)	(0.003)	(0.002)	(0.004)
Mother's education, 8–12 years	–0.006 **	–0.004	–0.001	0.003	0.001	0.004 *
	(0.003)	(0.004)	(0.004)	(0.002)	(0.001)	(0.002)
Mother's education > 12 years	–0.037 ***	–0.028 **	0.006	0.007	0.005	0.007
	(0.011)	(0.013)	(0.011)	(0.009)	(0.004)	(0.007)
Father's education, 8-12 years	#	#	#	#	#	#
Father's education > 12 years	#	#	#	#	#	#
N	*274581*	*274581*	*274581*	*274570*	*272838*	*274574*
Pseudo_R2	*0.11*	*0.02*	*0.04*	*0.04*	*0.03*	*0.02*

= dropped due to collinearity.

Single, double, and triple asterisks indicate significant coefficients at the 10%, 5% and 1% levels, respectively. The table shows marginal effects from probit estimations, confer eq. (1). Each column denotes separate regressions. Also included in the specifications are municipality and year-of-birth indicators. Standard errors are adjusted for clustering at the municipality level, and are available from the author.

Table 5. Heterogeneity in the response to the reform

Change in probabilities due to the educational reform. Timing of first birth and childlessness

	First birth at age 15-20			First birth at age 20-25			First birth at age 25-30			First birth at age 30-35			First birth at age 35-40			Being childless		
	Reform	Non-reform	Effect of reform	Reform	Non-reform	Effect of reform	Reform	Non-reform	Effect of reform	Reform	Non-reform	Effect of reform	Reform	Non-reform	Effect of reform	Reform	Non-reform	Effect of reform
Family income:																		
Bottom quartile	0.26	0.39	-0.12	0.37	0.42	-0.05	0.15	0.11	0.04	0.05	0.04	0.01	0.02	0.01	0.01	0.09	0.06	0.03
Above bottom quartile	0.11	0.08	0.03	0.40	0.37	0.03	0.24	0.26	-0.02	0.08	0.09	-0.01	0.03	0.03	0.00	0.10	0.12	-0.02
			-0.16			-0.08			0.07			0.02			0.01			0.05
Living in a major city:																		
Urban	0.12	0.18	-0.06	0.31	0.33	-0.01	0.25	0.22	0.04	0.10	0.09	0.02	0.02	0.02	0.00	0.12	0.09	0.03
Non-urban	0.11	0.12	-0.01	0.38	0.39	-0.01	0.23	0.22	0.02	0.07	0.07	0.00	0.03	0.02	0.00	0.12	0.10	0.02
			-0.05			0.00			0.02			0.01			0.00			0.01

Probabilities are computed after probit estimations, confer eq. (1). Family income is expressed through a dummy for whether or not the individual belonged to the lowest income quartile. When computing the probabilities, all variables except those specified in the table above (income, urbanity, reform) are kept at mean values. Also included in the specifications are each parent's age and level of education, as well as indicators for the woman's cohort and municipality. Standard errors are adjusted for clustering at the municipality level.

Table 6. Heterogeneity: income and urbanity combined

Change in probabilities due to reform. First birth at age 15–20

	Urban			Non-urban		
	Reform	Non-reform	Effect of reform	Reform	Non-reform	Effect of reform
Family income:						
Bottom quartile	0.27	0.47	–0.20	0.26	0.37	–0.11
Above bottom quartile	0.12	0.12	0.00	0.11	0.07	0.03
			–0.20			–0.15

Probabilities are computed after probit estimations, confer eq. (1). Family income is expressed through a dummy for whether or not the individual belonged to the lowest income quartile. When computing the probabilities, all variables except the specified (income, urbanity, reform) are kept at mean values. Also included in the specifications are each parent's age and level of education, as well as indicators for the woman's cohort and municipality. Standard errors are adjusted for clustering at the municipality level.

Table 7. Results, spillover effects among sisters

on probability of teenage motherhood

	Without background variables				With background variables			
	(I)		(II)		(III)		(IV)	
First birth at age 15–20	partial effect		partial effect		partial effect		partial effect	
Reform	–0.013	*	–0.014	**	–0.054	***	–0.054	***
	(0.007)		(0.007)		(0.016)		(0.016)	
1 if sister impacted by reform	0.012	*	–0.004		–0.006		–0.005	
	(0.007)		(0.008)		(0.006)		(0.007)	
Age difference between sisters			–0.006	***			0.000	
			(0.001)				(0.001)	
N	*48358*		*48358*		*48358*		*48358*	
Observed P	*0.174*		*0.174*		*0.174*		*0.174*	
Predicted P	*0.163*		*0.162*		*0.146*		*0.146*	

Single, double, and triple asterisks indicate significant coefficients at the 10%, 5% and 1% levels, respectively. The estimates show partial effects from probit models. Four different specifications have been used, which all relate to eq. (3): in (I) and (II), all arguments in the Z vector are set equal to zero, while the background variables Z are included in (III) and (IV). In specifications (I) and (III), the age difference variable D is omitted. Also included in each specification are municipality and year-of-birth indicators related to the younger sisters. Standard errors are adjusted for clustering at the municipality level.

Appendix

App. Table 1. Correlations

	Years of education	Reform	Year of birth	Urban	Mother's education, years	Mother's age	Father's education, years	Father's age	Family income	Mother's income	Father's income
Years of education	1.00										
Reform	0.09	1.00									
Year of birth	0.10	0.66	1.00								
Urban	0.08	0.09	–0.03	1.00							
Mother's education, years	0.37	0.07	0.08	0.15	1.00						
Mother's age	–0.01	–0.34	–0.51	0.03	–0.04	1.00					
Father's education, years	0.40	0.06	0.05	0.19	0.56	–0.01	1.00				
Father's age	–0.02	–0.32	–0.46	–0.01	–0.06	0.82	–0.03	1.00			
Family income	0.30	0.22	0.29	0.18	0.38	–0.20	0.41	–0.22	1.00		
Mother's income	0.13	0.07	0.07	0.14	0.26	–0.05	0.12	–0.05	0.41	1.00	
Father's income	0.28	0.22	0.29	0.14	0.32	–0.20	0.41	–0.23	0.94	0.09	1.00

In the table, "years of education" and "year of birth" refer to the 1947 to 1958 cohort member (n = 274,581).

App. Table 2. Summary statistics for the sister population

	Population of sisters			The whole sample		
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Explanatory variables:						
Years of education	91677	11.58	2.59	274581	11.52	2.57
Municipality	91677	1074.91	589.02	274581	1007.61	595.13
Reform	91677	0.57	0.50	274581	0.53	0.50
Year of birth	91677	1953.41	3.08	274581	1953.03	3.33
1 if born in 1958	91677	0.10	0.30	274581	0.10	0.29
1 if born in 1957	91677	0.10	0.29	274581	0.10	0.29
1 if born in 1956	91677	0.11	0.31	274581	0.10	0.30
1 if born in 1955	91677	0.11	0.31	274581	0.10	0.29
1 if born in 1954	91677	0.11	0.31	274581	0.09	0.29
1 if born in 1953	91677	0.11	0.31	274581	0.09	0.29
1 if born in 1952	91677	0.10	0.30	274581	0.09	0.28
1 if born in 1951	91677	0.08	0.27	274581	0.08	0.27
1 if born in 1950	91677	0.07	0.25	274581	0.07	0.26
1 if born in 1949	91677	0.05	0.23	274581	0.07	0.25
1 if born in 1948	91677	0.04	0.20	274581	0.06	0.24
1 if born in 1947	91677	0.04	0.18	274581	0.06	0.24
1 if lived in a major city	91677	0.14	0.35	274581	0.19	0.39
Outcome variables:						
1 if childless	91677	0.10	0.30	274581	0.11	0.31
1 if first birth at age 15–20	91677	0.16	0.37	274581	0.16	0.37
1 if first birth at age 20–25	91677	0.39	0.49	274581	0.39	0.49
1 if first birth at age 25–30	91677	0.23	0.42	274581	0.23	0.42
1 if first birth at age 30–35	91677	0.08	0.27	274581	0.08	0.27
1 if first birth at age 35–40	91677	0.03	0.16	274581	0.03	0.16

App. Table 3. Teenage motherhood by cohort

	Sister population (n = 91,677) Unconditioned			The whole sample (n = 274,581) Unconditioned		
Cohort	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
1947	3209	0.11	0.31	16156	0.14	0.35
1948	3954	0.12	0.33	17338	0.14	0.35
1949	4907	0.11	0.32	18499	0.14	0.34
1950	6310	0.11	0.31	20389	0.15	0.35
1951	7311	0.12	0.32	21552	0.16	0.37
1952	8860	0.16	0.37	24125	0.18	0.38
1953	9921	0.19	0.39	25667	0.19	0.39
1954	9940	0.20	0.40	25754	0.19	0.40
1955	9889	0.20	0.40	26091	0.18	0.39
1956	9721	0.18	0.39	26597	0.17	0.38
1957	8806	0.17	0.37	26179	0.15	0.36
1958	8849	0.16	0.37	26234	0.15	0.36

App. Table 4. Rank (reversed birth order)

Sister population (n = 91,677)

Rank within sisters in the family (1 = youngest)	Freq.	Per cent
1	43419	47.4
2	41197	44.9
3	6159	6.7
4	801	0.9
5	92	0.1
6	8	0.0
7	1	0.0

App. Table 5. Results, spillover effects among sisters

Teenage motherhood

	Without age difference:		With age difference:	
First birth at age 15–20	Partial effect	P>|z|	Partial effect	P>|z|
Reform	–0.054	0.00	–0.054	0.00
Age difference between sisters			0.000	0.72
Background variables:				
Urban	–0.008	0.56	–0.008	0.54
Family income, 2nd quartile	–0.139	0.00	–0.139	0.00
Family income, 3rd quartile	–0.144	0.00	–0.144	0.00
Family income, 4th quartile	–0.147	0.00	–0.147	0.00
Mother’s age, 1st quartile	0.089	0.00	0.090	0.00
Mother’s age, 2nd quartile	0.043	0.00	0.043	0.00
Mother’s age, 3rd quartile	0.031	0.00	0.031	0.00
Father’s age, 1st quartile	0.071	0.00	0.071	0.00
Father’s age, 2nd quartile	0.045	0.00	0.046	0.00
Father’s age, 3rd quartile	0.031	0.00	0.031	0.00
Mother’s education, 8–12 years	–0.058	0.00	–0.058	0.00
Mother’s education > 12 years	–0.082	0.00	–0.082	0.00
Father’s education, 8–12 years	–0.044	0.00	–0.044	0.00
Father’s education > 12 years	–0.095	0.00	–0.095	0.00
Interaction terms:				
1 if sister impacted by reform	–0.006	0.33	–0.005	0.48
Urban	0.003	0.87	0.003	0.88
Family income, 2nd quartile	0.114	0.00	0.114	0.00
Family income, 3rd quartile	0.115	0.00	0.115	0.00
Family income, 4th quartile	0.066	0.00	0.066	0.00
Mother’s age, 1st quartile	–0.016	0.31	–0.016	0.31
Mother’s age, 2nd quartile	–0.013	0.33	–0.013	0.33
Mother’s age, 3rd quartile	–0.020	0.11	–0.020	0.11
Father’s age, 1st quartile	–0.028	0.06	–0.028	0.06
Father’s age, 2nd quartile	–0.023	0.05	–0.023	0.05
Father’s age, 3rd quartile	–0.022	0.05	–0.022	0.05
Mother’s education, 8–12 years	0.002	0.87	0.002	0.87
Mother’s education > 12 years	–0.028	0.41	–0.029	0.41
N	*48358*		*48358*	
Observed P	*0.174*		*0.174*	
Predicted P	*0.146*		*0.146*	

The estimates show partial effects from probit models. The table reports results from two different specifications, which both relate to eq. (3), but in the second column the age difference variable *D* is left out. Also included in each specification are municipality and year-of-birth indicators related to the younger sisters. Standard errors are adjusted for clustering at the municipality level.