THREE ESSAYS ON THE ECONOMICS OF HEALTH HUMAN CAPTIAL AND HEALTH CARE

THREE ESSAYS ON THE ECONOMICS OF HEALTH HUMAN CAPTIAL AND HEALTH CARE

by

JINHU LI, B.A., M.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2011) McMaster University

(Economics) Hamilton, Ontario

Title:	Three Essays On the Economics of Health Human Capital and Health Care
Author:	Jinhu Li, B.A. (Huazhong University of Science and Technology), M.A. (The University of Victoria)
Supervisors:	Professor Jeremiah Hurley Professor Paul Contoyannis Professor Jeffery Racine
Number of Pages:	xiii, 266

Abstract

This thesis focuses on two important areas of health economics: health dynamics during pre-adulthood, and physician behaviour. Understanding the health development process of young people is of great importance for improving life-time well-being. The first two essays seek to explore the important factors that determine the health production process during the period of pre-adulthood. On the other hand, a better understanding of the behaviour of physicians, who are among the most important suppliers of health care, is of great importance for the design of social policies that aim to improve health of the population. The third chapter then turns the focus to physician labour and service provision behaviours.

The first chapter examines the impact of family social economic status (SES) and neighbourhood environment on the dynamics of child *physical* health development. It examines the distribution of health outcomes and health transitions and explores the determinants of these distributions by estimating the contributions of family SES, neighbourhood status, unobserved heterogeneity and pure state dependence.

The second chapter extends the research on health development in pre-adulthood by examining the roles of family SES, early childhood life-events, unobserved heterogeneity and pure state dependence in explaining the distribution of depression among adolescents and young adults. It also explicitly models the

depression dynamics and quantifies both the mobility and persistence of this type of *mental* health problem from adolescence to early adulthood.

The third chapter examines whether and how pay-for-performance (P4P) payments can motivate physician service provision to improve the quality of health care. It exploits a natural experiment in the province of Ontario, Canada to identify empirically the impact of P4P incentives on the provision of targeted primary care services, and whether physicians' responses differ by age, practice size and baseline compliance level.

Acknowledgements

First and foremost, I would like to thank my main thesis supervisor Professor Jeremiah Hurley. It was because of your constant guidance, encouragement and support that made it possible for me to complete my doctoral program and to become a health economist. I am very lucky to have you as my supervisor, to always be there to lead me through the whole process. You are a great mentor, and will be a life-time friend.

I would also like to thank my associate supervisor Professor Paul Contoyannis and thesis committee member Professor Jeffery Racine. This thesis would not be possible in its present form without your constant advice and guidance. I have benefitted immensely from your expertise, experience and your invaluable comments on each of the studies. I am truly fortunate to have worked with you both. In addition, I wish to thank Professor Phil DeCicca for being so kind to me-- you not only offered me your help and advice for my research, but also your support and friendship in the last three years of my doctoral program.

Special thanks to the generous financial support provided by the Research Program on the Social and Economic Dimensions of an Aging Population (SEDAP) for the last year of my PhD program, and throughout the years of my doctoral studies.

Thanks to my friends and classmates in the PhD program: Wei Yang, Simo

Goshev, Cong Li, Keqiang Hou, Daqing Luo, Kadir Atalay, Evan Meredith, Chao Wang and Taha Jamal. Thank you for discussing with me about research ideas, coming to my seminars, and helping me figure out how to use all kinds of statistical software. I have learned so much from you. It was because of your help and company that made my life a lot easier to go through the journey of my doctoral studies. My sincere appreciation to my other friends outside the PhD program: Jingjing Zhang, Emmanuel Guindon, Gioia Buckley, Huizi Zhao, James Chowhan and Junying Zhao. Your friendship truly doubled the joys and halved the sorrows in my life.

Finally, I would like to express my deepest gratitude to my dear family. I thank my mom Lanjiao Liu and my dad Sihai Li, for being the strongest support for my doctoral program, and for all the years of my education. I would never be able to finish my degrees without you. You have always been there with your patience, understanding and encouragement, to give me the strength to continue on with my study. Thank you for always believing in me, even when I have ceased to believe in myself. Because of your unconditional love, I become the person I am today. To my family, I dedicate this thesis.

Contents

List of Tables

List of Figures

Introduction

This thesis uses advanced micro-econometric techniques to investigate empirically a range of questions related to both health and health care. The research focuses on two important areas of health economics: health dynamics during pre-adulthood, and physician behaviour. The thesis consists of three essays. The first essay examines the impact of family social economic status (SES) and neighbourhood environment on the dynamics of child *physical* health development. The second essay examines the impact of family socio-economic conditions, stressful life-events and unobserved heterogeneity on the distribution and the dynamics of depression from adolescence to early adulthood. The third evaluates the impact of pay-for-performance incentive payments on the provision of health care by physicians.

The first two essays investigate health dynamics during childhood and adolescence. Health is an important form of human capital that influences social and economic success over the life cycle. Health development during the transition period of childhood to young adulthood is particularly important because initial adult health stock and attitudes towards health-promoting or risky behaviours are largely formed during this period (Heckman 2007). Moreover, evidence documents that pre-adult health is positively correlated with a spectrum

of health, educational and economic achievements over the lifespan (Case et al. 2005; Grossman 2000). Health may also play an important role in the intergenerational transmission of economic status (Currie 2009). Therefore, understanding the health development and the health production process of young people is of great importance for improving life-time well-being.

The classic model of the demand for health by Grossman describes the dynamic optimization problem of intertemporal utility maximization which leads to optimal lifetime paths of health capital and gross investment in health in each period (Grossman 1972; Grossman 2000). This model features the dynamic nature of the health production process. People experience persistently good or persistently poor health for two types of reasons: poor health can be inherently long-lasting and a cumulative history of health problems may have a direct effect on current health; disparities in factors like socioeconomic conditions may have long-lasting effects that influence health in multiple periods (Jones et al. 2006). Two of the thesis chapters seek to explore the important factors that determine the health development process and the dynamics of health during the period of childhood to early adulthood, and attempt to identify systematic differences in health persistence across subgroups of children and adolescents with heterogeneous family background or environmental characteristics. One chapter focuses on the roles of family socio-economic conditions and neighbourhood environments in determining the child *physical* health development, and the other explores the effects of family socio-economic conditions and childhood life

experience on one important aspect of youth *mental* health conditions-- depressive symptoms.

The first chapter uses data from the Canadian National Longitudinal Survey of Children and Youth (NLSCY) to investigate whether and why physical health outcomes exhibit persistence during the period from childhood to adolescence. On both efficiency and equity grounds, it is important to quantify the persistence of health over time and to identify systematic differences in persistence across different subgroups of the population. This chapter examines the distribution of health outcomes and health transitions using extensive descriptive analysis, and explores the determinants of these distributions by estimating the contributions of family SES, unobserved heterogeneity and pure state dependence while allowing for heterogeneity of state dependence parameters across categories of neighbourhood status.

The results of the dynamic models indicate a strong persistence of child physical health over time, and that certain community characteristics such as lower neighbourhood income, lower neighbourhood education level and higher proportion of lone-parents within a neighbourhood, contribute significantly to higher persistent levels of ill health over time. Moreover, the positive effect of "permanent" household income on child health is stronger in richer neighbourhoods and also more educated neighbourhoods, while the positive effect of "permanent" household income on child health is weaker in neighbourhoods with fewer lone-parent families and also in neighbourhoods with fewer families

living in rental accommodations. The chapter contributes to both the health dynamics literature and to the child health literature in two ways. First, few studies in the child health literature have been focused on modeling the evolution process of health outcomes from childhood to adolescence, particularly in Canada. Second, as this paper uses information on both family social economic status (SES) and neighbourhood level characteristics in a dynamic panel data framework, it contributes by examining the impact of contextual factors in the health dynamics literature. This paper has been published in the *Journal of Health Economics* (Contoyannis and Li 2011).

The second chapter extends the research on health development in pre-adulthood by examining, within both static and dynamic frameworks, the roles of family SES, early childhood life-events, unobserved heterogeneity and pure state dependence in explaining the distribution of depression among adolescents and young adults. Depression is one of the most common health problems in adolescence (Asarnow et al. 2009), and depression during this period often persists into adulthood and leads to adverse long-term outcomes (Colman et al. 2007; McLeod and Kaiser 2004). However, only a small number of empirical studies have examined the relationship between family or individual SES and depression among adolescents or young adults, and the results from these empirical studies are mixed. The first goal of this chapter is to examine the roles of family SES, early childhood life-events, and unobserved heterogeneity in explaining the distribution of depression among adolescents and young adults

using the US data on the children of the US National Longitudinal Survey of Youth 79 (NLSY79). Furthermore, this chapter explicitly models the depression dynamics and quantifies both the mobility and persistence of this type of mental health problem from adolescence to early adulthood, an issue that has not been addressed in other studies.

This study employs a conditional quantile regression framework. This is important because the factors of interest may not only affect the location of the conditional distribution of youth depression, but also affect the scale or other aspect of the distributional shape. If the underlying mechanism that links these factors with youth depression does differ at different parts of the depression distribution, using a conditional mean estimator will neglect this aspect and provide quite different policy implications. A methodological contribution of this chapter to the empirical health dynamics literature is that in addition to standard dynamic quantile regression models, it employs a newly-developed instrumental variable quantile regression for dynamic panel with fixed effects model to examine the dynamics of depression. This estimator not only allows us to control for individual-specific heterogeneity via fixed effects in the dynamic panel data framework, but also effectively reduces the dynamic bias generated by conventional dynamic fixed-effects estimation of the quantile regression models.

Results from the static conditional quantile estimation models reveal the asymmetry of the link between stressful life events and youth depression, and indicate the differential effects of family SES on youth depression at different

parts of the depression distribution. Results from the dynamic models suggest the importance of taking into account individual heterogeneity when examining the dynamics of youth depression. The results from the final instrumental variable with fixed effects model indicate that the pure state dependence of youth depression is very low and the observed positive association between previous depression and current depression is mainly due to unobserved individual heterogeneity.

One important determinant of the health production process is the consumption of health care. The Grossman model indicates that the demand for health care is a derived demand for health and is decided as part of the optimal choice over the gross investment in health stock over the life cycle (Grossman 1972; Grossman 2000). In a world full of asymmetric information, externalities and uncertainty, health care resource allocation via perfectly competitive markets leads to outcomes far from the Pareto optimum (Arrow 1963). Therefore, non-market institutions and public policies should step in to correct for market failures and to guide resource allocations for the improvement of social welfare (Hurley 2000). Given that the utilization of health care is largely determined by supplier behaviour, a better understanding of the behaviour of physicians, who are among the most important suppliers of health care, is of great importance for the design of social policies that aim to improve health of the population. Physicians' labour and service supply behaviours share some common features with workers in other industries, but are also distinguished by a greater influence of

professional standards and ethical concerns. In such a context, it is critical to conduct positive analysis that helps us to understand how contractual or institutional arrangements will affect their labour supply and service provision behaviours and facilitate the efficient and equitable allocation of health care resources. The third chapter of this thesis then turns the focus to physician labour and service provision behaviours.

The third chapter examines whether and how a certain type of financial incentives--- pay-for-performance payments --- can motivate physician service provision to improve the quality of health care. Explicit financial incentives, especially pay-for-performance (P4P) incentives, have been extensively employed in recent years by health plans and governments in an attempt to improve the quality of health care services. Classic principal-agent theory and incentive-contract theory suggest that performance-based contracting can induce agents to improve performance when payment is based on achieving pre-specified performance targets. However, using P4P programs to motivate health care providers' behaviour is controversial in reality. Theoretical predictions on physician responses to P4P incentives are ambiguous; there are still relatively few empirical studies that provide convincing evidence of how performance incentives influence physician delivery of targeted services. This chapter exploits a natural experiment in the province of Ontario, Canada to identify empirically the impact of P4P incentives on the provision of targeted primary care services, and whether physicians' responses differ by age, practice size and baseline compliance level.

The study uses administrative data that cover the full population of the province of Ontario and nearly all the services provided by practicing primary care physicians in Ontario. Different sources of administrative data are linked together to construct the individual-level data set of service provisions for physicians who were affected by the incentives (those work in the primary care reform models) and physicians that were not affected by the incentives (those work mainly in traditional fee-for-service practice) in both pre- and post-intervention periods. The study employs a difference-in-differences approach that controls for both selection on observables and selection on unobservables that may cause estimation bias in the identification. A set of robustness checks and sensitivity analyses are also implemented to control for potential confounding from other attributes of the primary care reform models, and from the other contemporary initiatives that could also influence the level of health care utilization during the study period.

The results indicate that, while all responses are of modest size, physicians responded to some of the financial incentives but not others. In general, the results confirm the empirical literature, which indicates little effect of employing P4P incentives to improve the quality of health care. The differential responses appear related to the cost of responding and the strength of the evidence linking a service with high-quality care, as well as the degree of complementarities between the P4P incentives and other institutional attributes of the practice models. Overall, the results provide a cautionary message regarding the effectiveness of

pay-for-performance schemes for increasing quality of care.

A common element across all three chapters is the application of advanced econometric techniques to help identify underlying causal relationships in areas of health economics where this has been a great challenge. In health economics, the prevalence of latent variables, unobserved heterogeneity and nonlinear dependent variables creates additional difficulties in the identification of causal relations. This thesis attempts to tackle various methodological issues in the identification strategies, and to select econometric tools that are well-suited for the data and for the estimation challenges encountered. The first two papers deal with the nonlinearity of the dependent variables by using latent variable models and count data models, and exploit longitudinal data sets to disentangle the pure state dependence from confounding by unobserved heterogeneity using a set of random effects and fixed effects models. The third paper exploits the exogenous variation in the primary variable of interest (the "treatment") generated by a quasi-experiment, and deals with the non-random selection and confounding problems by using a series of difference-in-differences models that gradually relax the exogeneity assumptions and by using multiple treatment or comparison groups.

References

Arrow, K. (1963), 'Uncertainty and the welfare economics of medical care', *American Economic Review* 53 (5): pp. 941–973.

Asarnow, J., McKowen, J., and Jaycox, L. (2009) 'Improving care for depression:

Integrating evidence-based depression treatment within primary care services', *Treatments for adolescent depression: Theory and practice.* (pp. 159-174). New York, NY, US: Oxford University Press. xii, 328 pp.

Case, A., Fertig, A. and Paxson, C. (2005) 'The lasting impact of childhood health and circumstance', *Journal of Health Economics* 24: 365-389

Colman, I., Wadsworth, M., Croudace, T., and Jones, P. (2007) 'Forty-year psychiatric outcomes following assessment for internalizing disorder in adolescence', *The American Journal of Psychiatry*, Vol 164(1), pp. 126-133

Contoyannis, P. and Li, J. (2011) 'The evolution of health outcomes from childhood to adolescence', *Journal of Health Economics*, Volume 30, Issue 1, Pages 11-32

Currie, J. (2009) 'Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development', *Journal of Economic Literature,* 47:1,87-122

Grossman, M. (1972) 'On the concept of health capital and the demand for health', *Journal of Political Economy* 80(2), 223-55.

Grossman, M. (2000) 'The Human Capital Model', In *The Handbook of Health Economics*, ed. Anthony Culyer and Joseph P. Newhouse (Amsterdam: North Holland)

Heckman, J. (2007) 'The Economics, Technology and Neuroscience of Human Capability Formation', *Proceedings of the National Academy of Sciences*, 104(33): 13250-13255.

Hurley, J.(2000) 'An overview of the normative economics of the health sector', Chapter 2 in the *Handbook of Health Economics*, Volume 1, Part 1, 2000, Pages 55-118

Jones, A.M., Rice, N. and Contoyannis, P. (2006) 'The dynamics of health', Chapter 2 in *The Elgar Companion To Health Economics*, Edward Elgar Publishing Limited.

McLeod, J., and Kaiser, K., (2004) 'Childhood emotional and behavioural problems and educational attainment', *American sociological review*, vol. 69, no5, pp. 636-658

Chapter 1
The Evolution of Health Outcomes from Childhood to Adolescence[1]

1.1 Introduction

Health development during the period of childhood to adolescence is important because, for most individuals, initial health in adulthood and attitudes towards health promoting or risky behaviors are largely formed during this transition period (Heckman 2007). Furthermore, evidence documents that pre-adult health is positively correlated with achievement over the lifespan (see e.g. Case et al 2005). While the association of child/youth health and economic, institutional and environmental factors has been examined by various studies within a static framework, few studies have focused explicitly on health dynamics from childhood to adolescence. On both efficiency and equity grounds, it is important to quantify both the mobility and persistence of health over time and to identify systematic differences in mobility across subgroups. Knowing the systematic differences in the dynamics of health across different subgroups helps to

[1] A slightly revised version of this chapter was published in the Journal of Health Economics: Contoyannis, P. and Li, J. (2011) 'The evolution of health outcomes from childhood to adolescence', *Journal of Health Economics*, Volume 30, Issue 1, Pages 11-32.

disentangle how different factors determine the health transition from childhood to adolescence within a population. Furthermore, if we observe that reductions in health status are more permanent than transitory in nature for particular groups, we may be more concerned about this than cross-sectional variation in health; more efficient improvement of average health status of the whole population can be made possible if social support programs are targeted at individuals who are more likely to have multiple periods of ill-health and equity objectives likely require us to be more concerned about children who suffer prolonged ill health.

This study draws on two streams of health outcomes research. The first stream focuses on the association of child/youth health and economic, institutional and contextual factors. A positive relationship between high family SES and good child health status has been recorded in various studies. Using cross-sectional data sets of U.S children, Case et al. (2002) pointed out children's health is positively related to household income and the income-health gradient has deepened as children age. They also investigated the extent to which the gradient can be explained by other characteristics of children and parents, including child health at birth, parental health, genetic ties, health insurance and maternal labor supply. Following Case et al. (2002), Currie and Stabile (2003) used the Canadian NLSCY to confirm the deepening gradient, and to test two hypotheses of the underlying mechanisms that cause the deepening gradient. They concluded that the mechanism of the deepening gradient is not that children with poorer health lack the resources to respond to health shocks, but they are subject to more shocks.

Curtis et al. (2001) explored data from the Ontario Child Health Study (OCHS) to estimate the association between child health and both low-income and family status. They find that child health is much more strongly (and negatively) related to low-average-income than to low-current income, while lone-mother status is negatively associated with child outcomes. Contoyannis and Dooley (2010) examined the relationship between childhood health problems and various young adult outcomes and the role that health status plays in the intergenerational correlation of economic outcomes using the Ontario Child Health Study (OCHS). Specifically, they examined the association between parental socio-economic status and the prevalence of a childhood chronic condition, a functional limitation, or a conduct or emotional disorder and reported for each case an income effect that is modest in size. They also found that parental health is strongly related to childhood health outcomes, but the effect of family income on child health is not mainly a proxy for parental health. Another two studies have provided evidence of the health-SES gradients among adolescents (Graeme Fort et al. 1994, Chris Power and Sharon Matthews 1997). The above examples largely identified the potential SES factors that are correlated with and may contribute to the health of children and adolescents. However, it is worth mentioning that few of these studies are implemented in a panel data framework and dealt with individual unobserved heterogeneity. The only study we are aware of which involves the transition of health outcomes from childhood to adolescence is Currie and Stabile (2003). In order to test one of the two hypotheses in explaining the deepening SES-health

gradient recorded by Case et al. (2002), they investigated whether low-SES children deal with bad "health shocks" as effectively as high-SES children by examining if the negative impacts of previous chronic conditions onset differ by family SES. While their results are in line with ours in the sense that poor health status in the previous period has persistent negative effects on current child health, the study did not focus on how state dependence systematically determine the dynamics of child health over time and how state dependence of child health differ across neighborhood types as in our study. In their study, only two periods of data are used and the onset of chronic conditions in the first period are controlled as the "health shocks" for health state in the second period; while in our study all six cycles are used and self-assessed health status in the previous period is controlled for in modeling current self-assessed health status.

The positive association between SES and health is difficult to untangle for adults, due to the likelihood of a reverse causal relationship. Although the channel that runs directly from health to income can be eliminated for the case of children, possible unobserved factors that can affect child health outcomes and are also correlated with family SES make identification of a causal relationship difficult. Dooley and Stewart (2004) used data from the Canadian NLSCY and cautiously estimated the size of the effect of income on child's cognitive outcomes by attempting to separate out the variation in outcomes caused by potential unobserved heterogeneity and that caused by regressors. They implemented four empirical strategies using panel data and reported a smaller income effect on child

outcomes than from conventional estimates which are obtained from weighted least squares regressions with pooled data. This difference in estimates reveals the benefit of exploiting a panel data structure when unobserved individual heterogeneity contributes substantially to child outcomes.

Other studies have focused on the social contextual influences on child outcomes. Boyle et al. (2007) used multilevel models to examine longitudinal associations between contextual influences (neighborhood and family) and educational attainment in a cohort of 2,355 children. The results showed that while 33.64% of the variation in individual level educational attainment can be explained by their model, 14.53% of the variation is attributable to neighborhood and family-level variables versus 10.94% to child-level variables. Several other studies have provided consistent evidence that neighborhood or community level socioeconomic advantage is positively associated with better child outcomes (Brooks-Gunn, Duncan, Klebanov and Sealand 1993; Garner and Raudenbush 1991). Leventhal and Brooks-Gunn (2000) provide a comprehensive review of research on the effects of neighborhood residence on child and adolescent well-being. By summarizing the existing evidence of neighborhood effects on child and youth outcomes, they conclude that high SES is of great importance for school readiness and achievement while low SES and residential instability are determinants of poor behavioral/emotional outcomes. Therefore, social contextual or environmental characteristics should be considered as other important factors related to child and youth health.

The second stream of studies on health outcomes focuses on modeling adult health distributions in a dynamic framework. Studies have addressed the question of why some adults experience persistently good or bad health. The persistence could be explained by pure state dependence, particular individual socio-economic characteristics, or environmental characteristics (Jones, Rice and Contoyannis 2006). Some empirical health dynamics studies have examined the relative contributions of pure state dependence and unobserved heterogeneity, and the conditional effect of socio-economic status in explaining observed health status variation (Contoyannis, Jones and Rice 2004a, Contoyannis, Jones and Rice 2004b), while other empirical health dynamics studies have provided evidence of associations between observed health *persistence* and SES positions. In particular using the British Household Panel Survey (BHPS), Hauck and Rice (2004) found evidence of substantial mental health mobility and that the extent of mobility varies across SES categories with greatest persistence in lower income groups and less educated individuals. In a different framework, Buckley et al. (2004) examined the influence of SES position on transition probabilities from good health to poor health for older Canadians. The results showed that the probability of remaining in good health is higher in the highest quartile of income and education, which also indicated a positive association between good health and SES.

Our study aims to contribute in the following ways. Firstly, this study contributes to the health dynamics and child health literature. As discussed above

few studies have been focused on modeling the evolution process of health outcomes from childhood to adolescence, particularly in Canada. Secondly, as this paper uses information on both family SES positions and neighborhood level characteristics into the dynamic panel data framework, it contributes by examining the impact of contextual factors in the health dynamics literature.

This paper proceeds as follows. Section 2 describes the data set we used for the study and presents some descriptive analysis of the data. Section 3 introduces the theoretical rationale and empirical framework of the study. In section 4, the regression results are reported and analysed while in section 5 some conclusions are provided.

1.2 Data

As this study considers both the effects of family SES positions and neighbourhood characteristics on child health dynamics, two data sets are explored in our study. The first data set is the Canadian National Longitudinal Survey of Children and Youth (NLSCY) cycles 1 to 6, which contains rich information on child outcomes and family SES positions. The second data set is the Census profile data of Canada 1996 and 2001, which contains information on neighborhood characteristics. We construct and use the following four sets of variables throughout this study: 1) child general physical health outcome measures, e.g. Self-Assessed Health (SAH) of the child reported by the Person

Most Knowledgeable(PMK) about this child; 2) family socio-economic variables, e.g. total household income, parental education, family structure (family size, whether the child is living with two parents) etc.; 3) Other variables for the child and the parents such as age, whether the PMK is the biological parent of the child and maternal age at birth of the child; 4) neighborhood level variables, indicating the "affluence" status and "socioeconomic disadvantage" status of the neighborhoods, e.g. mean household income, percentage of population with university degree, etc.

1.2.1 Sample and variables

The National Longitudinal Survey of Children and Youth (NLSCY) is the main data source used in this study to examine the contribution of individual and family level variables in determining health transitions. The NLSCY is a survey "designed to collect detailed information every two years about the factors influencing a child's cognitive, emotional and physical development and to monitor the impact of these factors over time" (NLSCY user guide). With the main purpose of following up a group of children over time, the survey began to collect information with one large cohort of 0-11 year- olds in 1994, and followed up every two years till 2004 (Cycle 6). All the available waves so far (from Cycle 1 to Cycle 6) are used in this study.

As stated in the NLSCY User's Guide, the NLSCY is divided into four components: the household component, adult component, child component, and

youth component. The household component is used to determine the relationship between all household members. It also identifies the person most knowledgeable (PMK) about the child in the household. The PMK provides the information for all selected children in the household and then gives information about himself/herself and his/her spouse/partner. A child component was created for each selected child between 0 and 17 years of age. The PMK about children and youth answered the child component questions. The child component provides information on the child demographic information and child health measures. But the only sections of the Child Questionnaire asked about youth aged 16 and 17 are the Aspirations and Expectations section, Custody and the Socio-Demographics section. Therefore, the relevant child health information is available in this component only for children aged 15 and younger. We could find health measures for the children/youth aged 16 and older in the Youth Questionnaire, as the youth component is used for selected respondents aged 16 to 21 years old. However, the respondent of the Youth Questionnaire answer questions about themselves so we suspect that the reporting would be systematically different from the responses from the PMKs. An adult component was created for the PMK and his/her spouse or partner, if the selected child is 17 years old or younger. This component collects information for the PMK and the spouse of the PMK about their age, education, income, labor force participation and health condition etc. From this information, the family structure and parental characteristics with potential impacts on child's health development are extracted.

With respect to child health, the variable of general health assessed by the PMK is used in the analysis. The survey question requires the respondent to rank the child's health as excellent, very good, good, fair or poor. This measure falls into the category of a subjective measure of self-assessed health (SAH) which provides ordinal rankings of the respondents' perceived health status. Although the reliability of this subjective measure of health has been questioned by some literature (see Crossley and Kennedy 2002), the child health measure is confined to this variable in our study for the following reasons. Firstly, measures of self-assessed health are commonly used in the literature and have generally been found to be powerful predictors of mortality (see Idler and Kasl 1995; Idler and Benyamini1997; Burström and Fredlund 2001), and to be good predictors of subsequent use of medical care (see van Doorslaer et al. 2000, 2002). Also, since SAH has been consistently defined across different datasets based on which most empirical studies are conducted, using the same measure makes our results more comparable to the others. The study from Crossley and Kennedy (2002) has provided evidence that this measure suffers from the non-random measurement error in terms of reporting, and the perceptions of the respondents' own health systematically vary by age and some socioeconomic status. However, our study is limited by the availability of other suitable measures of health[2]. Other concerns

[2] The McMaster Health Utility Index Mark 3 (HUI3) is often deemed a more objective measure of general health but this measure is only available for children aged 4 or 5 years old in the NLSCY. Other existing measures of self-reported chronic conditions in NLSCY do not provide us a global measure of general health of children. It is worth noting that even the self-reported objective measures of health on the incidence of chronic conditions are criticized for the significant

about this measure are related to the reporting heterogeneity in the ordered responses which may invalidate group comparisons and measures of health inequality (Lindeboom and van Doorslaer 2004; Murray et. al 2001). More objective measures of health are suggested and methods to overcome this problem are discussed in this literature (see discussion in Contoyannis, Jones and Rice 2004b).

In order to investigate the relationship between family SES and child health outcomes we use the total household income in the past 12 months and a set of variables for parental educational achievements. Case et al. 2002 found that while there still exists a large and significant correlation between income and child's health, the addition of parental education levels to the regression controls had a substantial impact on the estimated income coefficients (reducing the magnitude of the positive correlation). This suggests that household income and parental education are two important factors in determining the child's health and they affect child's health through different pathways. In the NLSCY, information about educational attainment, labor force participation etc. are collected for the PMK and the spouse of PMK, but the PMK and the spouse of PMK are not necessarily the biological parents of the child. They can be step parents, adopting parents or even unrelated persons. This brings in complexity in interpretation because mother's education may influence child health through both her childcare skills after birth and the health of the child at birth, while a PMK who is not the mother

---

measurement error. Details see Michael Baker, Mark Stabile and Catherine Deri 2004.

will likely exert a much larger influence(relative to the birth mother) on child health through childcare. Moreover, mother's education and father's education level are expected to have different impacts on child's health in that, in most cases, it is the mother who takes care of the child and their behavior would shape child's health to a larger extent, especially for the children at younger ages. Therefore, we separate mother's education from father's education level. In this study, mother's education was obtained from the PMK's (or the spouse of PMK) education level if PMK (or the spouse of PMK) is the biological mother of the child. Otherwise, female caregiver education was obtained from the closest female figure in the household (defining the biological mother as the closest female figure overall), i.e. it was obtained from the information of the PMK (or the spouse of PMK) if PMK (or the spouse of PMK) is female but not the biological mother of the child. If there is no education information for the closest adult female figure in the household, female caregiver education was set to missing. The variable for male caregiver education was derived in the same way. In order to capture the difference between the effects of education for a biological mother and another female figure, a dummy indicating the PMK (or spouse of the PMK) is the biological mother of the child is included in the regression and interacted with mother's education level. Also, a dummy indicating PMK is female is included in the regression to account for the response "bias" by gender. Other than the main SES variables, family structure characteristics have a potential impact on child health. A variable for family size indicating the total number of persons living in

the household and a dummy variable indicating whether or not a child lives with both parents are included in the regression too[3]. Table 1 in the Appendix A lists the definitions of the main variables we used in this study.

To explore the relationship between neighborhood characteristics and child health dynamics, we split our sample by a set of neighborhood level variables indicating the "affluence" status and "socioeconomic disadvantage" status of the neighborhood the child resides in. In our study, "neighborhood" is defined by census tract (CT) boundaries within all census metropolitan areas (CMAs) and part of census agglomerations (CAs) where a CT boundary exists, while by Enumeration Area (EA) or dissemination areas (DAs) boundaries within more rural areas where a CT boundary does not exist. Census tracts (CTs) are small geographic units representing urban or rural neighborhood-like communities within all CMAs and CAs with an urban core population of 50,000 or more at the previous census. In most CTs, there are 2,500-8,000 people living within them (Statistics Canada, 1992). An EA is the smallest level of geographical aggregation used by Statistics Canada: it contains at least 375 dwellings in urban areas and 125 dwellings in rural areas. To attach neighborhood information to every child in each cycle, we firstly matched the neighborhoods identities within NLSCY and Census profile data through Enumeration Area (EA) or Dissemination Area (DA) code which exist in both data sets. Since the neighborhoods are mostly defined by

[3] For the same reasons, we discriminate between the scenario of "child living with both biological parents" and the scenario of "child living with both parental figures but not the biological parents" and included both variables in our analyses.

CT boundaries, we then used the Geography Tape File (GTF) to map from EA/DA boundaries to CT boundaries when CT boundaries are used to define neighborhoods. At the end, the neighborhood variables aggregated at the CT boundary level are used for the neighborhoods defined by CTs; while the neighborhood variables aggregated at the EA or DA boundary level are used for the neighborhoods defined by EAs or DA s. In our study, the "affluence" status of the neighborhoods is measured by two variables: average household income and the percentage of the adult population with university or college degrees; while the "socioeconomic disadvantage" status of the neighborhoods is measured by another two variables: percentage of families headed by lone parents and the percentage of families living in rental accommodations. These specific concepts of community characteristics have been established and used in studies examining the neighborhood influence on educational attainment of children (Boyle et al. 2007). Since we are using a longitudinal cohort and the respondents might have moved from one neighborhood to another across cycles, we mapped the respondents into neighborhoods for each cycle based on the most up-to-date available census profile data at that time. In other words, the neighborhood characteristics are drawn from the census profile data 1996 for the first four cycles of NLSCY, while these values are drawn from the census profile data 2001 for the last two cycles of NLSCY.

1.2.2 Data description

As we focus on the longitudinal transition of the child health distribution over time, our study employs data on the original longitudinal cohort in NLSCY over six waves. There is considerable attrition in the longitudinal cohort of the NLSCY. According to the NLSCY Cycle 7 User Guide, by cycle 6, "the cumulative, longitudinal response rate for children in the original cohort was 57.6%"[4]. Because of the sample attrition, around 11,000 children aged 10 to 21 years old from the original longitudinal cohort remained in the sample. Several sample selection criteria have been used for the investigation of family SES and child health dynamics association in our study. Firstly, we only included children aged 0-15 (including age 15) in all cycles. As discussed earlier, in the NLSCY the self-assessed general health (SAH) status is reported in the Child Questionnaire by the Person Most Knowledgeable (PMK) about the child for children aged 0 to 15; while this health measure is reported in the Youth Questionnaire by the youth themselves for children aged 16 and older. The PMK is an adult figure, usually the mother/father of the child. We believe the response from the PMKs and the response from the children themselves are systematically different so we excluded the children aged 16 and older. This leads to a reduction of our study sample to

[4] In order to adjust for total non-response, the NLSCY employs weighting procedures to produce two longitudinal (funnel and non-funnel) weights at each cycle. Specifically, these weights are calculated by taking the child's design weight and making adjustments for survey non-response and post-stratification to ensure that the final survey weights sum to known counts of children by age, sex and province (See the NLSCY User's Guide in references) for the attrition rate and the weighting procedure which attempts to adjust for total non-response). Accordingly, we applied the funnel weights to our final sample in the descriptive analysis because funnel weights are assigned to children who have responded at every cycle.

6,611 children. Secondly, we only included children who had information with respect to all of our main variables listed in Table 1 in Appendix A in all six cycles. In other words, only a balanced panel sample is used for both descriptive and regression analysis. This leads to a further reduction of the study sample to 3,752 children. Thirdly, we excluded children with obvious errors in their data, e.g. we excluded children who had multiple gender values across cycles. We ended up with 22,398 observations for 3,733 children with 6 time periods as our study sample. For the subgroup analysis with different neighbourhood status, we then only included children with complete information with respect to the four neighbourhood variables in all six cycles. This leads to a further reduction of sample to 21,726 observations for 3,621 children with 6 time periods[5].

1.2.2.1 The study sample

Child SAH

Originally the health status variable is a categorical variable with 5 ranks. However, we regrouped this variable in the descriptive analysis by merging the fair health group and poor health group because of the constraint imposed by the data confidentiality requirement from Statistics Canada[6]. After the merge, the

[5] We lost 112 children from our analyses for neighborhood effects on child health dynamics because a) some of the EA or DA codes are missing from the NLSCY; or b) some of the EA or DA codes of our NLSCY sample cannot be found in the Census profile data; or c) at least one of the four neighborhood variables are missing values in the corresponding Census profile data.

[6] Statistics Canada's data confidentiality restriction requires that—"Data users must not release or publish any estimate that would allow the identification of a specific respondent or reveal any individual's responses. For this reason, estimates (for example, the cells in a cross-tabulation) should have at least five contributing respondents" (NLSCY cycle 7 User's Guide). As only a small proportion of children in our sample reported poor health in all cycles, we had to regroup the

number of observations in the fair/poor health group is big enough for data disclosure. Figure 1 (see all figures in Appendix A) shows the health dynamics of children over 6 cycles. The proportion of children in excellent health was decreasing and the proportion of children in very good health was increasing slightly between cycles 1 and 3. Between cycles 4 and 6 there does not appear to be a discernible trend in the proportions reporting excellent and very good health. In all cycles there are only a very small proportion of children reported as in fair or poor health with no apparent trend in this proportion or for the proportion in good health.

Figure 2 displays the distribution of child's health status pooled over 6 cycles by household income categories. From the figure, it can be seen that children's health status is better in households with higher incomes than those in households with lower incomes. As we move from low income group to high income groups, the proportion of children in excellent health increases while the proportion of children in fair or poor health decreases.

Figures 3 displays the distribution of child's health status pooled over 6 cycles, by mother's education attainment. The figure shows very similar patterns of child health variation as to household income level. The proportion of children with excellent health increases and the proportion of children with fair or poor health decreases as we move up from lower maternal education level to higher maternal education level. The pattern can be observed as well in the distribution

---

two categories of "poor" and "fair" health together to reach the minimum cell size.

of child's health by father's education attainment.

State Dependence

State dependence in health has been explored by the literature on health dynamics (e.g Contoyannis et al. 2004) and it is expected to explain a substantial proportion of health variation. Without conditioning on other variables, the degree of mobility/persistence of health outcomes can be assessed descriptively by the probability distribution conditioned on the previous health distribution. Figure 4 shows the distribution of child's health status in cycle 2 by the previous health status in cycle 1. It can be seen from the figure that given the child was in excellent health in cycle 1, the probability of transiting from excellent health to fair or poor health is very close to zero and the probability of staying in excellent health is very high. Similarly, for the children who had fair or poor health in cycle 1, the probability of transiting from fair or poor health to excellent health is very low while the probability of staying fair or poor health is high. In general, this figure shows that children are much more likely to stay in their health status of origin than moving away from it. The same pattern can be seen for all the cycles from a transition matrix in Table 1. The elements of the table can be interpreted as the conditional probabilities under a Markov model. The table shows that conditioning on being in excellent or very good health states, children are much more likely to stay within the states than moving away from them in the current period; while conditioning on being in good health or lower than good health, children are more likely to move one level up in the current period. It indicates

that the persistence mainly operates around the state of excellent health and very good health while the health status is pretty mobile around the states of good and fair/poor health.

Family SES and other variables

In order to examine the association between family SES characteristics and child health dynamics, we compared the means of the family SES variables across a set of child health transition scenarios. Tables 2 and 3 present the means for the main family SES and other demographic variables for the study sample and for a set of interesting sub-samples by health transition patterns. Column 1 in Table 2 lists the mean values for the whole balanced sample. The second column shows the average characteristics for the children who had excellent or very good health for all 6 cycles and the third column shows the average characteristics for the children who always had less than good health. Column 4 presents the mean values for the children who had a single transition from excellent or very good health to worse health status without recovering to the original health status, while column 5 shows the mean of variables for the children who had a single transition from less than good health to better health and stayed healthy since then. From the comparison between the second and third columns, it can be seen that children who were always in excellent or very good health tend to be living in a smaller household and be brought up in a richer family than the children who were always in good or less than good health. Also, mother's age at the birth of the child is lower for the children with excellent health or very good health than for the

children with good or less than good health. Surprisingly, there is no systematic difference in the parents' education level for these subgroups. No specific pattern is found comparing the subgroup of children who had a single transition from excellent to very good health and did not recover and the subgroup of children who had a single transition from good to poor health, except that household income and parents' education level are slightly higher for the first subgroup than for the second subgroup.

In Table 3, we show the mean values of these variables for the subsample of children who had few health drops[7] versus the subsample of children who had multiple drops, and for the subsample of children whose health drop lasted for only 1 cycle versus the subsample of children whose health drop lasted for multiple periods. Columns 1-4 show the mean values for the groups of children who had 0, 1, 2, 3 or 4 drops during our study period, respectively. Children with lower household income and lower parental education tend to experience multiple health drops relative to the children with higher family SES. This observation is in line with the result from the study by Currie and Stabile (2003) which indicates that children brought up in families with lower SES are subject to more health shocks than the children with higher family SES. Columns 5-8 show the mean values for the groups of children who had 1 drop and this drop lasted for only 1cycle, for 2 cycles, for 3 cycles and for 4 cycles. A slight negative association is discernable from the comparison among these neighborhood subsamples, with

[7] A "health drop" here is defined as a decrease of SAH from any health status (e.g. excellent to fair or very good to poor). The decrease could be 1 category or more.

children who experienced short health drops are brought up in families with slightly higher income. The basic descriptive statistics shows a negative association between family SES and the number of health shocks the children experienced while a much weaker negative association exists between the family SES variables and the persistence of health shocks.

1.2.2.2 Sub-samples by long-term neighbourhood status

State Dependence

Another goal of this study is to identify which neighbourhood characteristics contribute to the persistence of poor health states. To examine the heterogeneity of the state dependence across neighbourhood characteristics, we divide the study sample into four subgroups for each of the four neighbourhood variables and constructed the transition matrices for each subgroup[8]. When we split the sample into subgroups, we divide them into quartiles based on the simple average of a neighborhood variable across 6 cycles. This allows us to include both movers and stayers in our study sample and does not restrict classification according to the neighborhood variable at an arbitrary period of time for all individuals (e.g cycle 1). We can see some general patterns over a set of transition matrices presented in Table 4. The first panel of Table 4 shows the transition matrices for neighborhoods with lowest, second lowest, middle and highest levels of average household

[8] Here we regrouped the five ranks into three ranks because of the confidentiality restriction from Statistics Canada noted above. We combined poor, fair and good health into a category of "equal or less than good health" so that for all cross-tabs the cell size is greater than 5. Accordingly, we have only three categories of health status for the descriptive statistics in the subgroup analysis: excellent, very good and equal or less than good health.

income, respectively. It shows that the less than good health state is more persistent in lower income neighborhoods than in higher income neighborhoods. In particular, in the highest income neighborhoods children with less than good health in the last period are most likely to move up one rank, while in the lowest income neighborhoods they are most likely to continue to have less than good health. The second panel shows the transition matrices for neighborhoods with less educated people and for neighborhoods with more educated people. Being in the less than good health state is more persistent in neighborhoods with less educated people than in neighborhoods with more educated people. The third panel presents the transition matrices for neighborhoods with larger proportions of families headed by lone-parents and for neighborhoods with smaller proportions of families head by lone-parents. The last panel shows the transition matrices for neighborhoods with larger proportions of families living in rental accommodations with smaller proportions of families living in rental accommodations. The similar pattern in these four panels indicates that, without conditioning on any other family-level variables, the persistence level of ill health is different across neighborhoods with different socio-economic conditions. In particular, the ill health state is more mobile in neighborhoods with higher income, in neighborhoods with more educated people, in neighborhoods with fewer families headed by lone-parents and in neighborhoods with fewer families living in rental accommodations.

1.3 Empirical Methods

A widely used economic model (Currie 2000) for child health determination will be followed in this study. In the standard model, parents are assumed to maximize an inter-temporal utility function, which trades off child's health stock and their consumption of other goods and leisure, subject to a series of budget and time constraints. The solution to the maximization problem gives the demand function for child health stock. Unfortunately we do not know the health production function which makes it impossible to specify the complete structural model and, in any case, it is difficult to estimate convincingly. Therefore, an alternative representation is used instead in which child health outcomes depend on a set of family SES factors (mainly family income, family structure), child characteristics, parental characteristics and some initial conditions such as maternal age at birth.

Empirically, this study will examine the effects on child health outcomes of SES position, neighbourhood characteristics, pure state dependence and unobserved heterogeneity. Taking into account neighbourhood characteristics is expected to reduce estimates of unobserved heterogeneity. State dependence will be taken into account by controlling for the lag of the health status of the child, while unobserved heterogeneity will be controlled for by using random effects models. Previous empirical studies have been implemented using either pooled approaches or dynamic nonlinear panel data approach with random effects (Contoyannis et al. 2004a, b, Hauck and Rice 2004). This is because, with a nonlinear fixed effect model, the MLE estimator is not consistent in a panel

setting with small T (# of time periods) and large N (# of individuals), due to the incidental parameters problem from estimating the fixed effects.

As in most of the micro-level panel data cases, our data is a short panel of large cross-sections (large N but small T). Econometricians have attempted to find fixed-T consistent estimators in modelling discrete choices with individual effects but, in general, fixed-T consistent estimators for nonlinear panel models are not available for most models with unobserved heterogeneity treated as fixed effects. As in static models, there is a trade-off between choosing fixed and random effects approaches for the dynamic nonlinear panel data models we consider in this study, in the sense that achieving fixed-T identification with a less restricted conditional distribution of individual effects usually requires a more restrictive specification of the conditional distribution for y given variables of interest and individual effects (e.g. logit type).

Fixed effects models are more robust without imposing restrictions on the conditional distribution of individual effects but it suffers from the incidental parameter problem. There are no general solutions for nonlinear models with fixed effects, and in some cases, although a specific solution is available, it is not root-N-consistent. For example a dynamic logit fixed T- consistent estimator is available but it converges slowly and does not allow for time dummies. (see Honore and Tamer 2006).

Arellano (2003) pointed out that there are random effects models that achieve fixed T consistency subject to a particular specification of the form of the

dependence between the explanatory variables and the effects, but they rely on strong and untestable auxiliary assumptions. For example, the random effects dynamic nonlinear panel data approach advocated by Woodridge (2005), which is one of the approaches we implement in our study, can generate consistent estimators only when the specified distribution of the individual effects is correct. Even though fixed T consistency is achievable for less restrictive random effects specifications, identification is often out of reach (see Honore and Tamer 2006).

1.3.1 Baseline dynamic panel ordered probit model without individual effects

A basic approach to estimating the effect of family SES variables in explaining the health transition is to estimate a dynamic panel model without dealing with individual specific effects at all. We denote this the pooled model. The regression model can be simply specified as below:

$$H_{it}^{*} = \theta' H_{it-1} + \beta' X_{it} + \varepsilon_{it} \quad \text{(i=1,…, N; t=2, …, T)}, \qquad (1)$$

where H_{it}^{*} is the latent variable of health outcome, H_{it-1} is a vector of indicators for the child's health status in the previous period, X_{it} is a set of observed family SES variables. ε_{it} is a time and individual-specific error term which is assumed to be normally distributed and uncorrelated across individuals and waves. The latent variable H_{it}^{*} relates to the observed health outcome H_{it}

as follows:

$$H_{it} = j \quad \text{if} \quad \mu_{j-1} < H_{it}^* < \mu_j, j = 1,\ldots,m \quad , \qquad (2)$$

where $\mu_0 = -\infty, \mu_j \leq \mu_{j+1}, \mu_m = \infty$.

1.3.2 Dynamic panel ordered probit model with random effects

The empirical specification incorporating the family SES effect and unobserved heterogeneity can be written as:

$$H_{it}^* = \theta' H_{it-1} + \beta' X_{it} + \alpha_i + \varepsilon_{it} \quad \text{(i=1, …, N; t=2, …, T)} , \qquad (3)$$

where α_i is an individual-specific and time-invariant random component, and the idiosyncratic component ε_{it} is assumed to be uncorrelated with α_i. The latent variable H_{it}^* specification is the same as in 3.1.1.

This study follows the approach of Wooldridge (Wooldridge 2005), Contoyannis et al. (2004b) which attempts to deal with the initial conditions problem in non-linear dynamic random effects models; the individual specific effect is specified as the following:

$$\alpha_i = \alpha_0 + \alpha_1' H_{i1} + \alpha_2' \overline{X}_i + u_i \quad , \qquad (4)$$

where $\overline{X}_i$ is the average over the sample period of the observations on the time-varying exogenous variables and u_i is assumed to be normally distributed. When the error process is not serially independent and the initial observations are not the true initial outcome of the process thus are not exogenous in nature, treating

the lagged dependent variables as exogenous leads to inconsistent estimators in non-linear dynamic random effects models. Equation 4) deals with this initial conditions problem by directly modeling the distribution of the unobserved effect as a function of the initial value and any exogenous explanatory variables. However, as discussed earlier, since this approach specifies a complete model for the unobserved effects, the consistency of the estimator can be sensitive to misspecification of this distribution.

1.4 Estimation Results

1.4.1 Family SES and child health distribution

We explore the determinants of child health distributions by estimating the contributions of family SES, unobserved heterogeneity and state dependence with the dynamic panel data models described in the previous section. Table 5 presents the coefficient estimates for the ordered probit models based on pooled and random effects specifications. Column 1 and 2 shows the estimates of coefficients and standard errors with the pooled ordered probit model, while column 3 and 4 show the estimates of coefficients and standard errors with the random effects model with the specification suggested by Wooldridge (2005). The pooled ordered probit models allow for serial correlation in the errors by using a robust estimator of the covariance matrix. Several patterns can be seen from the comparison of the models. Firstly, there is a gradient in the effect of previous

health on current health. The reference group here is the group reporting very good health (the second highest rank of health state). For both of the models, previous health is highly statistically significant and the magnitude of the coefficient is not trivial. Secondly, the child's health status does improve as family SES position increases, shown by the significant and positive coefficients on the household income variable and positive gradients on parental education level. In order to capture the differential effects of maternal education on child health through biological and other pre and postnatal effects, the interaction terms of maternal education with the dummy indicating whether the PMK is the biological mother of the child are included in the regressions. It can be seen from column 3 and column 4 that after controlling for the within-individual average of current household income and the within-individual average of parental education level, and with adjustments for unobserved heterogeneity in the estimation procedure, the original current household income variable and parental education variables are not as large and some are no longer statistically significant. This result is in line with the interpretation of regarding the mean income as a measure of long-term or 'permanent' income while regarding current income as a measure of transitory income shocks (Contoyannis et al. 2004 a, b). It shows that the long-term household income, other than the transitory income, is important for the child's health status. Other statistically significant variables are child age, and age of mother at birth of child, and family size. Thirdly, the improvement in the log-likelihood from the pooled model to the random-effects model indicates that

allowing for unobserved heterogeneity can improve the goodness-of-fit of the model. Moreover, it can be seen from the ICC value in the random-effects model that about 31% of the latent error variance is attributable to unobserved heterogeneity.

As the estimated coefficients for the pooled models are not directly comparable to the ones for the random effects models, we calculated the average partial effects (APEs) on the probability of reporting excellent health. Following the approach of Wooldridge (Wooldridge 2005), Contoyannis et al. (2004b), we calculated the average partial effects (APEs) by computing the partial effect at the observed values of the regressors for each observation and averaging the estimates over all the observations [9]. The results are presented in Table 6. The random-effects model results indicate that, relative to the children who reported very good health in the previous period, the children who reported excellent health in the previous period are more likely to stay in excellent health in the current period by 9.12 percentage points, while the children who reported good health, fair health and poor health previously are less likely to report excellent health in the current period by 7.23 percentage points, by 13.57 percentage points and by 27.12 percentage points, respectively.

An "empirical" transition matrix of reporting each health status given the

[9] As usual, the partial effects are obtained by taking the derivative of the ordered probit probabilities with respect to the variable in question for continuous regressors; while for discrete regressors, they are obtained by taking differences. Wooldridge (2005) shows that computing the partial effect at the observed values of the regressors for each observation and averaging the estimates over the observations provides a consistent estimate of the APE.

previous health status is constructed based on the estimates of the random effects model and reported in Table 7. The way we construct the empirical transition matrix is as follows. First, the probabilities of reporting each health state are predicted and generated for each observation based on the estimated parameters. Second, all the observations are pooled together and grouped by the previous health status. For each of these groups, the means of the predicted probabilities of reporting each health status are calculated and these constitute the point estimates of the transitional probabilities. This transition matrix is comparable to Table 1 except that it shows the predicted transitional probabilities conditional on all the family-level control variables. The elements on the diagonal of Table 7 are smaller than the ones of Table 1. This highlights the importance of family-level characteristics and unobserved individual effects in explaining the persistence of child health status over time.

1.4.2 Neighbourhood characteristics and child health transitions

1.4.2.1 Long-term neighbourhood characteristics and child health transitions

As in the descriptive analyses, we divide the study sample into quartiles based on the simple average across 6 cycles of each of the four neighborhood variables: average household income of the neighbourhood, the proportion of the population with a college degree, the proportion of families headed by lone-parents and the proportion of households living in rental accommodation. Since these measures

are essentially the within-means of neighbourhood characteristics for each child, they can be interpreted as the long-term neighbourhood environment rather than the temporary neighbourhood characteristics. For each neighbourhood subsample, we estimated a pooled ordered probit model and random effect ordered probit model with the specification suggested by Wooldridge (2005). The corresponding average partial effects (APEs) of reporting excellent health status for the random effects specification are presented in Part A of Table 8a, 8b, 8c and 8d for each of the four neighbourhood characteristics. The gradient of pure state dependence is observable across all neighbourhood subsamples. "Permanent" household income has significant positive effects on reporting excellent health for all the subgroups, but the magnitudes of the effects indicate different interaction patterns between "permanent" household income and different neighbourhood characteristics. For example, the positive effect of "permanent" household income on child health is stronger in richer neighbourhoods and also more educated neighbourhoods. This shows the average household income level and education level of neighbourhood are positive moderators of a "permanent" family income effect. On the contrary, the positive effect of "permanent" household income on child health is weaker in neighbourhoods with less lone-parents families and also in neighbourhoods with less families living in rental accommodations. Maternal education has significant positive effects on reporting excellent health for most of the subgroups, while the neighbourhood characteristics have negative moderating effects on the effect of maternal education. Maternal education plays a more important role in the most

disadvantaged neighbourhoods relative to better neighbourhoods. No discernable pattern can be found for the effect of paternal education on child health distributions.

To illustrate that living in different types of neighborhood leads to significantly different health dynamics in the long term, we conducted a one-to-one comparison on the average partial effects (APE) of each health lag term across all neighborhood quartiles, and we implemented a simple test that examines whether each pair of the APE estimates are significantly different. The test-statistic and the p-values are presented in Part B of Table 8a, 8b, 8c and 8d for each of the four neighbourhood characteristics. The results from the tests confirm that the persistence level differs systematically across different neighborhood status except for neighborhood living arrangements.

A set of empirical transition matrices of reporting each health status given the previous health status for different types of neighbourhoods are constructed based on the estimates of the random effects model and reported in Table 9. These transition matrices are comparable to the ones in the descriptive analysis except that they are the predicted probabilities conditional on all the control variables. In the table, previous health status is presented in rows while current health status is presented in columns. Like the transition matrices in the descriptive analysis, the low health state is more persistent in neighborhoods with lower income, in neighborhoods with less educated people and in neighborhoods with more families headed by lone-parents than in neighborhoods with better conditions.

Nonetheless, there is no discernable pattern across neighborhoods with different living arrangements defined by the proportion of families living in rental accommodations. It indicates that controlling for family level characteristics neighborhood income, neighborhood education and neighborhood lone-parents status remain important in explaining the heterogeneity of persistence levels of ill-health over time.

In order to show the magnitude of the difference in the transition probabilities across different neighborhood quartiles, we constructed 95% confidence intervals for each estimate of the transition probabilities for all empirical transition matrices in Table 9. The point estimates and the 95% confidence intervals of the transitional probabilities[10] are presented by figure 5 to figure 8 for each of the four neighborhood characteristics. These figures illustrate which neighborhood characteristics contribute to the difference in the dynamics and to what extent the transition probabilities differ across quartiles by these neighborhood characteristics. From the figures, we see that the difference in health transitions across quartiles of neighborhood income is most obvious. Five out of the nine transitions of health status have systematically different transitional probabilities across neighborhood income quartiles. Differences in health transitions are also discernable across quartiles of neighborhood education and quartiles of

[10] There are 9 types of transitions in our case here: transition from "<=Good health" to "<=Good health", transition from "<=Good health" to "Very Good health", transition from "<=Good health" to "Excellent health", transition from "Very Good health" to "<=Good health", transition from "Very Good health" to "Very Good health", transition from "Very Good health" to "Excellent health", transition from "Excellent health" to "<=Good health", transition from "Excellent health" to "Very Good health" and transition from "Excellent health" to "Excellent health".

neighborhood lone-parents status, but in only three out of the nine transitions. No difference in health transitions is observed across quartiles of neighborhood living arrangements. Overall, the transitional probability of being in "less than good health" and stuck in this poor health status in the next period is systematically lower in richer neighborhoods, neighborhoods with more educated people and in neighborhoods with fewer families headed by lone-parents. The transitional probability of being in excellent health and staying in excellent health in the next period is systematically higher in richer neighborhoods, neighborhoods with more educated people and in neighborhoods with fewer families headed by lone-parents.

Furthermore, we calculated the predicted probabilities of trajectories of some specific health transition scenarios[11] based on these transition matrices. Figure 9 shows the predicted probabilities of health drops lasting for only 1 period versus health drops lasting for multiple periods across different neighbourhoods. The first panel compares the probabilities across neighbourhoods with different levels of average household income. The second, third and fourth panel compares the probabilities across neighbourhoods with different proportions of highly-educated people, across neighbourhoods with different proportions of lone-parents families and across neighbourhoods with different proportions of families living in rental accommodations. Figure 10 shows the predicted probabilities of children having 0 drop, 1 drop, 2 drops, 3 or 4 drops during 6 cycles across different

[11] These health transition scenarios correspond to the ones listed in table 2 and table 3.

neighbourhoods. It is observable that children tend to experience multiple health drops living in poorer neighborhoods, in neighborhoods with less educated people, in neighborhoods with more families headed by lone-parents and in neighborhoods with more families living in rental accommodations.

To test if there is any effect of current neighborhood characteristics on child health dynamics, we also estimated the same pooled ordered probit and random effects ordered probit models with our full sample on an alternative specification which includes interaction terms between the health lags and the concurrent neighborhood variables[12]. The regression results show that the gradient in the estimated effect of previous health on current health (estimated coefficients of the health lag dummies) are still clear and significant, while most of the estimated coefficients of the interaction terms are insignificant. In order to test the hypotheses that (at least some) current neighborhood characteristics do affect the transition dynamics of child health, we conducted a Wald test on the joint significance of each set of the interaction terms, e.g. interactions terms between health lag dummies, initial health status and the neighborhood income quartiles. The results from the Wald tests indicate that current neighborhood characteristics in general do not moderate the transition dynamics in a significant way, except for

[12] We constructed the interaction terms for our random effect model as follows. For each neighborhood characteristic (i.e. income, education, lone-parents status and living arrangement) the neighborhood quartiles are now constructed according to the quartile the child was in during each cycle. The four sets of neighborhood quartile dummies (4*3=12 more regressors in total) are included in the regression along with the interactions with the health lag dummies (3*3*4= 36 more regressors), plus the interactions with the initial health state dummies (3*3*4= 36 more regressors). So in this regression, we are using the full sample instead of the subsamples while estimating 84 additional parameters in the model.

neighborhood living arrangement condition. In summary, the regression results from this model indicate that the concurrent neighborhood characteristics do not have a significant impact on child health transitions, or that the impact (if there is indeed an impact) couldn't be detected by the random-effects model using our study sample. Given that our previous subgroup analyses by different quartiles of average neighborhood characteristics had different persistence level in health dynamics, we conclude that it is the long-term neighborhood/environmental conditions (other than concurrent conditions) that are contributing to the difference in the child health transition.

1.4.2.2 Neighbourhood transition paths and child health

One might argue that not only the average environment characteristics for the children could affect the transitions of child health but also particular types of change in the environment over time could lead to very different dynamics. To explore the potential effects of the change in the environment on the dynamics of child health, we conducted subgroup analyses based on different "transition paths" of neighborhood characteristics[13]. At first we assigned the neighborhood quartile of each of the four neighborhood variables in every cycle to each child. As a result,

[13] A criticism of using simple averages of neighborhood variables to divide the sample is that it might not capture the effect of neighborhood changes on the dynamics of child health if there are a lot of changes in the environment over time for the children and these changes lead to different dynamics. Now we looked further into the "transition paths" of neighborhood and examine if we could test this hypothesis. This test is feasible because there is sufficient variation in terms of the neighborhood changes in our sample: around 43% of the children in our sample stayed within the same neighborhood income quartile over 6 cycles, about 36% of the children moved once from one quartile to another, while about 21% of them moved twice or more across different neighborhood income quartiles. These percentage figures are similar in terms of the movement across other neighborhood characteristics.

every child has a sequence of environmental positions over 6 cycles. Then we group the sample based on the direction of these "transition paths": moving to better neighborhoods over time ("climbing-up" pattern), moving to worse neighborhoods over time ("sliding-down" pattern), moving to better neighborhoods at one time then moving to worse neighborhoods at another ("bouncing" pattern), or staying in the same type of neighborhood over time[14]. Using these neighborhood subsamples, we estimated the pooled ordered probit model and random effects ordered probit model and again compared the estimated state dependence parameters to examine if there is any different dynamics across different transition paths of environment.

With the subgroup of children who stayed in the same neighborhood quartile over time, we split them into four subgroups by quartiles of neighborhood status they stayed in over six cycles and again constructed the empirical transition matrices among these four groups. Table 10 presents the empirical transition matrices by quartiles of neighborhood characteristics among these children who didn't change their neighborhood status over six cycles. Similar to the pattern showed in Table 9, the low health state is more persistent in neighborhoods with lower income, in neighborhoods with less educated people and in neighborhoods with more families headed by lone-parents than in neighborhoods with better

[14] Because there are too many different "transition paths" over the six cycles according to the way we sliced our sample, (e.g. being in the highest income quartile for the first 3 cycles while moving to the second lowest quartile for the next 3 cycles), we were not able to estimate our random effects model on each subsample (due to sample size restrictions).

conditions, while no discernable pattern is found across neighborhoods with different living arrangements. The fact that the same pattern is preserved among the "stayers" indicates the results are robust to different study samples.

Table 11 presents the empirical transition matrices by neighborhood transition patterns. In general, all the transitional probabilities are very similar across different neighborhood transition patterns, indicating that there is no significant impact of neighborhood transition patterns on the dynamics of child health.

1.5 Conclusion

We explored the relative contributions of family SES, unobserved heterogeneity and state dependence in determining child health distributions. From the descriptive analysis, the positive correlation between SES and child health can be seen: children in household with higher income and more educated parents tend to be healthier in general. The results from the regression analysis indicate that the child's health status does improve as family SES position increases with household income having a large and positive effect on child health. However, after adding in the mean household income into the regression, the current household income is no longer statistically significant and the coefficient of mean household income shows a positive impact of long-term income on child health. The same pattern is found for parental education. Positive state dependence of child health is observed from the results in all dynamic models. The coefficients

of health lags indicate persistence in health from childhood to adolescence. Using Wooldridge's random effects specification, unobserved heterogeneity explained approximately 31% of the latent error variance.

We also examined the potential effects of neighborhood contextual factors on the dynamics of child health by estimating the dynamic panel data models allowing for heterogeneity of state dependence parameters across categories of neighborhood status. The regression results from the subgroup analyses indicate that the positive effect of "permanent" household income on child health is stronger in richer neighbourhoods and also more educated neighbourhoods, while the positive effect of "permanent" household income on child health is weaker in neighbourhoods with fewer lone-parents families and also in neighbourhoods with fewer families living in rental accommodations. Taken together, this may highlight one of the important mechanisms through which neighbourhood contextual factors can influence child outcomes-- collective efficacy serves as a key neighborhood process likely to impact on developmental health (Sampson, Raudenbush and Earls 1997). In other words, the social exchanges of residents in richer neighborhoods and more educated neighborhoods could lead to a more efficient process which magnifies the protective effect of family income in the production of child health. The persistence level differs systematically across different neighborhood status except for neighborhood living arrangements. Specifically, the poor health status is more persistent in neighborhoods with lower income, in neighborhoods with less educated people and in neighborhoods with

more families headed by lone-parents than in neighborhoods with better conditions. Results from alternative models indicate that it is the long-term neighborhood or environmental conditions, other than temporal conditions that are contributing to the difference in the child health transition. Furthermore, transition patterns of neighborhood characteristics do not explain the variability of child health dynamics over time. Accordingly, the predictions from the analyses based on long-term neighborhood status indicate that children living in poorer neighborhoods and in neighborhoods with lower education level tend to experience poor health status for longer after a transition to it, while children tend to experience multiple health drops living in poorer neighborhoods, in neighborhoods with less educated people, in neighborhoods with more families headed by lone-parents and in neighborhoods with more families living in rental accommodations.

Our study suffers from several limitations. First, our estimation results may suffer from potential bias generated by partial non-response of the NLSCY, as we are only using a balanced-sample in our study. Children are dropped from our study when some of the family-level SES measures are missing from the data. Second, the four variables we chose as measures of neighborhood characteristics might not be sufficiently comprehensive to capture the contextual factors that are important in determining the dynamics of child health. Besides neighborhood "status" indicators, characteristics representing the "capacity" and the "process" of neighborhoods are also identified as important contextual factors in child health

development (Leventhal and Brooks-Gunn 2000; Sampson, Morenoff and Gannon-Rowley 2002). Our study could be extended to examine the effect of more neighborhood characteristics including the quality of institutional resources and public infrastructure in the neighborhoods and neighborhood collective efficacy. Third, the random effects model we employed in our analyses can generate consistent estimators only when the specified distribution of the individual effects is correct. Fixed effects estimation is more robust than random effects estimation as it avoids the initial conditions problem and the specification of the relationship between the individual effects and regressors in the model, although it suffers from the incidental parameter problem. There are no general solutions for nonlinear models with fixed effects, and in some cases, although a specific solution is available, it is not root-N-consistent. A literature has been specifically focused on bias-adjusted methods of estimation of nonlinear panel data models with fixed effects. One future extension of our study is to employ a Modified Maximum Likelihood Estimation (MMLE) approach that reduces the order of the score bias from $O(T^{-1})$ to $O(T^{-2})$ regardless of the existence of an information orthogonal re-parameterization (Carro 2007) to provide more robust empirical results.

References

Arellano, M.(2003) 'Discrete choice with panel data'. *Investigaciones Economicas* XXУIII (3), 423-458

Arellano, M.(2005) 'The Cox-Reid modified score: a comment' Addendum of 'Discrete choice with panel data'

Arellano, M. and Hahn, J. (2005) 'Understanding Bias in Nonlinear Panel Models: Some Recent Developments,' Working Papers wp2005_0507, CEMFI.

Baker, Michael, Mark Stabile, and Catherine Deri. (2004) 'What Do Self-Reported, Objective, Measures of Health Measure?' *Journal of Human Resources* 39(4): 1067-1093.

Boyle, M., Georgiades, K., Racine, Y. and Mustard, C. (2007) 'Neighbourhood and family influences on educational attainment: results from the Ontario Child Health Study follow-up 2001' *Child Development* vol. 78, number 1: 168-189

Brooks-Gunn, J., Duncan, G., Klebanov, P.K., and Sealand, N. (1993) 'Do neighbourhoods influence child and adolescent development?' *American Journal of Sociology* Vol. 99: 353-395

Buckley, N.J., Denton, F. T., Robb, A.L. and Spencer, B. G. (2004) 'The transition from good to poor health: an econometric study of the old population' *Journal of Health Economics* vol. 23: 1013-1034

Burström B, Fredlund P. (2001) 'Self rated health: is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes?' *Journal of Epidemiology and Community Health* 55: 836-840.

Canadian Institute for Health Information (CIHI) (2005) 'Improving the health of Young Canadians' Ottawa: Author. <http://www.cihi.ca/youth>

Carro, J.M., (2007) 'Estimating dynamic panel data discrete choice models with fixed effects'. *Journal of Econometrics* Vol. 140: 503-528

Case, A., Lubotsky, D. and Paxson, C. (2002) 'Economic Status and Health in Childhood: The Origins of the Gradient' *American Economic Review* Vol. 92, 5: 1308-1334

Case, A., Fertig, A and Paxson, C. (2005) 'The lasting impact of childhood health and circumstance' *Journal of Health Economics* 24: 365-389

Contoyannis, P. and Dooley, M. (2010) 'The role of child health and economic status in educational, health, and labour market outcomes in young adulthood'

Canadian Journal of Economics, 43: 323–346.

Contoyannis, P., Jones, A.M., and Rice, N. (2004a) 'Simulation-based inference in dynamic panel ordered probit models: an application to health' *Empirical Economics*, vol. 29: 49-77

Contoyannis, P., Jones, A.M., and Rice, N. (2004b) 'The dynamics of health in the British Household Panel Survey' *Journal of Applied Econometrics*, vol. 19: 473-503

Cox, D.R. and Reid, N. (1987) 'Parameter orthogonality and approximate conditional inference' *Journal of the Royal Statistics Society*, Series B 49: 1-39

Crossley, Thomas F. & Kennedy, Steven (2002) 'The reliability of self-assessed health status,' *Journal of Health Economics*, Elsevier, vol. 21(4), pages 643-658, July.

Currie, J (2000) 'Child health in developed countries' in Joseph P. Newhouse and Anthony J. Culyer, eds., *Handbook of health economics*, Vol. 17. Amsterdam: North-Holland, pp. 1053-90

Currie, J and Stabile, M. (2003) 'Socio-economic status and child health: Why is the relationship stronger for older children?' *American Economic Review* Vol. 93, 5: 1813-1823

Curtis, Lori J.; Dooley, Martin; Lipman, Ellen L. and Feeny, David H. (2001) 'The role of permanent income and family structure in the determination of child health in Canada' *Health Economics*, Vol. 10 (4): 287-302

Deaton, Angus and Paxson, Christina (1999) 'Mortality, education, income and inequality among American cohorts' *NBER working paper* No. 7141

Dooley, Martin and Stewart, Jennifer (2004) 'Family income and child outcomes in Canada' *Canadian Journal of Economics*, Vol. 37(4): 898-917

Ford, Graeme et al. (1994) 'Patterns of class inequality in health through the life-span: class gradients at 15, 35, and 55 years in the west of Scotland' *Social Science and Medicine*, 39(8): 1037-50

Garner, C. and Raudenbush, S.W. (1991) 'Neighbourhood effects on educational attainment: A multilevel analysis' *Sociology of Education*, Vol.6: 251-262

Hauck, K. and Rice, N. (2004) 'A longitudinal analysis of mental health mobility in Britain', *Health Economics*, vol. 13: 981-1001

Heckman, James J. (2007) 'The Economics, Technology, and Neuroscience of

Human Capability Formation' *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 104, No. 33 (Aug. 14, 2007): 13250-13255

Honore, B.E. and Tamer, E. (2006) 'Bounds on parameters in dynamic discrete choice models', *Econometrica*, vol. 74: 611-629

Idler EL, Benyamini Y(1997) 'Self-rated health and mortality: a review of twenty-seven community studies'. *Journal of Health and Social Behavior* 38(1): 21-37.

Idler EL, Kasl SV. (1995) 'Self-ratings of health: do they also predict change in functional ability?' *Journal of Gerontology* 50B: S344-S353.

Jones, A.M., Rice, N. and Contoyannis, P. (2006) 'The dynamics of health', Chapter 2 in *The Elgar Companion To Health Economics*, Edward Elgar Publishing Limited.

Leventhal, T. and Brooks-Gunn, J. (2000) 'The neighbourhoods they live in: the effects of neighbourhood residence on child and adolescent outcomes' *Psychological Bulletin*, vol. 126, no. 2: 309-337

Lindeboom, Maarten & van Doorslaer, Eddy (2004) 'Cut-point shift and index shift in self-reported health,' *Journal of Health Economics*, Elsevier, vol. 23(6), pages 1083-1099

Murray CJL, Tandon A, Salomon J, Mathers CD (2001) 'Enhancing cross-population comparability of survey results'. GPE Discussion Paper No. 35, WHO/EIP, Geneva.

National Longitudinal Survey of Children and Youth (NLSCY), Cycle 7 - Microdata User Guide. Statistics Canada

Power, C. and Matthews, S. (1997) 'Origins of health inequalities in a national population sample' *Lancet*, 350(9091): 1584-1589

Sampson, R., Morenoff, J., Gannon-Rowley, T. (2002) 'Assessing neighbourhood effects: social process and new directions of research'. *Annual Review of Sociology*, 28, 443-478

Sampson, R., Raudenbush, S., Earls, F. (1997) 'Neighbourhoods and violent crime: a multilevel study of collective efficacy'. *Science*, 277, 918-924

van Doorslaer E et al. (2000) 'Equity in the delivery of health care in Europe and the US'. *Journal of Health Economics* 19: 553-583.

van Doorslaer E, Jones AM, Koolman X. (2002) 'Explaining income-related

inequalities in doctor utilization in Europe: a decomposition approach'. ECuity II Project, Working Paper #5.

Wooldridge, J. (2005) 'Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity' *Journal of Applied Econometrics* vol. 20: 39-54

Table 1.1: Transition matrix, balanced study sample

		Fair/Poor t	Good t	Very Good t	Excellent t
Fair/Poor	t-1	0.250	0.411	0.199	0.140
Good	t-1	0.043	0.355	0.378	0.224
Very Good	t-1	0.010	0.124	0.460	0.405
Excellent	t-1	0.005	0.042	0.219	0.735

Table 1.2: Mean of family SES and other variables

	(1)	(2)	(3)	(4)	(5)
Variables	Whole balanced sample	Always in excellent or very good health	Always less than good health	Single transition from excellent or very good health to worse health	Single transition from less than good health to better health
	N=22,398	N=14,676	N=120	N=15,870	N=1,416
child age	7.480	7.429	7.039	7.477	7.429
child gender	0.492	0.483	0.422	0.478	0.582
family size	4.512	4.538	5.570	4.525	4.502
mother's age	29.346	29.626	31.842	29.534	29.055
household	71,125.0	75,395.8	49,355.5	73,833.9	70,115.0
schoolm1	0.092	0.070	0.099	0.074	0.116
schoolm2	0.220	0.209	0.088	0.219	0.185
schoolm3	0.212	0.212	0.307	0.211	0.254
schoolm4	0.475	0.509	0.507	0.495	0.445
schoolf1	0.131	0.113	0.139	0.118	0.163
schoolf2	0.216	0.210	0.157	0.217	0.239
schoolf3	0.189	0.187	0.324	0.186	0.205
schoolf4	0.464	0.491	0.381	0.480	0.394
PMK not	0.074	0.079	NA	0.079	0.062
PMK female	0.928	0.922	NA	0.923	0.952
Living w/	0.988	0.991	NA	0.990	0.994

1. schoolm1, schoolm2, schoolm3 and schoolm4 are the percentages of female caregivers whose highest education is less than secondary, equal to secondary school graduation, some post-secondary and college or university degree, respectively.

2. schoolf1, schoolf2, schoolf3 and schoolf4 are the percentages of male caregivers whose highest education is less than secondary, equal to secondary school graduation, some post-secondary and college or university degree, respectively.

3. NA=Not available due to Statistics Canada Research Data Centre restrictions[15].

[15] According to Statistics Canada Research Data Center (RDC) program guidelines, with the NLSCY only statistics based on greater than 5 observations can be released outside of RDCs.

Table 1.3: Mean of family SES and other variables

Variables	(1) Had 0 drop	(2) Had 1 drop	(3) Had 2 drops	(4) Had 3 or 4 drops	(5) Had 1 drop & duration =1 cycle	(6) Had 1 drop & duration =2 cycles	(7) Had 1 drop & duration =3 cycles	(8) Had 1 drop & duration =4 cycles
	N=6,480	N=9,768	N=5,370	N=780	N=3,888	N=1,248	N=522	N=174
child age	7.532	7.442	7.467	7.628	7.300	7.385	7.069	7.155
child gender	0.477	0.493	0.501	0.543	0.550	0.492	0.428	0.530
family size	4.601	4.462	4.473	4.656	4.496	4.383	4.326	4.455
mother's age at	29.656	29.355	29.027	28.777	29.262	29.170	30.575	31.351
household income	81,648.8	69,959.0	61,824.1	59,616.4	71,718.5	68,052.7	66,493.1	67,001.3
schoolm1	0.058	0.095	0.118	0.185	0.078	0.077	0.052	0.254
schoolm2	0.196	0.217	0.254	0.221	0.186	0.193	0.181	0.247
schoolm3	0.225	0.211	0.198	0.212	0.214	0.210	0.300	0.159
schoolm4	0.520	0.477	0.430	0.382	0.523	0.520	0.467	0.340
schoolf1	0.094	0.127	0.173	0.213	0.124	0.110	0.087	0.254
schoolf2	0.206	0.218	0.234	0.152	0.193	0.250	0.169	0.180
schoolf3	0.179	0.205	0.172	0.187	0.199	0.211	0.282	0.137
schoolf4	0.521	0.450	0.421	0.448	0.484	0.428	0.462	0.429
PMK not mother	0.085	0.069	0.072	0.059	0.070	0.048	0.060	0.043
PMK female	0.921	0.931	0.928	0.941	0.930	0.953	0.940	0.957
Living w/ both	0.990	0.995	0.978	0.965	0.996	0.986	NA	NA

Table 1.4: Descriptive transition matrices by long-term neighborhood status

By quartiles of mean household income of neighbourhood																
Lowest income				Second lowest income				Middle income				Highest income				
	<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex	
<=Good	0.470	0.363	0.166	<=Good	0.455	0.309	0.236	<=Good	0.481	0.285	0.234	<=Good	0.322	0.459	0.220	
Very	0.133	0.469	0.398	Very	0.133	0.486	0.381	Very	0.149	0.453	0.398	Very	0.127	0.443	0.430	
Excellent	0.064	0.278	0.658	Excellent	0.050	0.229	0.721	Excellent	0.053	0.236	0.711	Excellent	0.034	0.176	0.790	
By quartiles of proportion of population with university degree in neighbourhood																
Lowest % with college degree				Second lowest %				Second highest %				Highest % with college degree				
	<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex	
<=Good	0.480	0.317	0.202	<=Good	0.419	0.349	0.232	<=Good	0.455	0.330	0.216	<=Good	0.346	0.443	0.212	
Very	0.148	0.454	0.398	Very	0.160	0.413	0.427	Very	0.126	0.462	0.412	Very	0.118	0.494	0.388	
Excellent	0.058	0.237	0.705	Excellent	0.057	0.240	0.703	Excellent	0.042	0.232	0.726	Excellent	0.039	0.184	0.777	
By quartiles of proportion of families headed by lone-parents in neighborhood																
Highest % with lone-parents				Second highest %				Second lowest %				Lowest % with lone-parents				
	<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex	
<=Good	0.503	0.304	0.193	<=Good	0.395	0.377	0.229	<=Good	0.392	0.385	0.223	<=Good	0.365	0.402	0.233	
Very	0.149	0.466	0.386	Very	0.133	0.443	0.424	Very	0.102	0.469	0.428	Very	0.157	0.462	0.381	
Excellent	0.061	0.270	0.669	Excellent	0.046	0.196	0.758	Excellent	0.039	0.198	0.763	Excellent	0.040	0.207	0.753	
By quartiles of proportion of families living in rental accommodations in neighborhood																
Highest % with rental accommodations				Second highest %				Second lowest %				Lowest % with rental accommodations				
	<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex		<=Good	Very	Ex	
<=Good	0.469	0.323	0.208	<=Good	0.441	0.364	0.196	<=Good	0.408	0.350	0.242	<=Good	0.356	0.423	0.221	
Very	0.134	0.477	0.389	Very	0.148	0.420	0.431	Very	0.135	0.458	0.408	Very	0.121	0.483	0.396	
Excellent	0.058	0.242	0.700	Excellent	0.039	0.215	0.746	Excellent	0.046	0.208	0.746	Excellent	0.042	0.199	0.759	

Table 1.5: Dynamic ordered probit models estimates

	(1)	(2)	(3)	(4)
	Pooled model, without correlated		Random effects, with correlated	
hlthc(t-1)poor	-1.9473	(0.2692)	-0.9073	(0.2619)
hlthc(t-1)fair	-1.1681	(0.0941)	-0.4432	(0.0913)
hlthc(t-1)good	-0.5473	(0.0328)	-0.2361	(0.0357)
hlthc(t-1)excellent	0.7523	(0.0218)	0.2963	(0.0266)
child age	-0.0054	(0.0028)	-0.0030	(0.0040)
child gender	-0.0415	(0.0195)	-0.0492	(0.0291)
family size	0.0292	(0.0099)	-0.0345	(0.0270)
mbirthage	-0.0077	(0.0023)	-0.0159	(0.0035)
ln(hh income)	0.1718	(0.0208)	0.0247	(0.0353)
mother school2	0.1349	(0.0376)	0.1035	(0.0550)
mother school3	0.1694	(0.0392)	0.0904	(0.0655)
mother school4	0.2255	(0.0372)	0.1250	(0.0725)
father school2	0.0676	(0.0323)	0.0146	(0.0457)
father school3	0.0578	(0.0343)	-0.0441	(0.0568)
father school4	0.0697	(0.0312)	-0.0823	(0.0634)
PMK not mother	-0.5410	(0.3185)	-0.4902	(0.7221)
mother	-0.2598	(0.1695)	-0.3210	(0.1838)
mother	-0.1644	(0.1578)	-0.1568	(0.1889)
mother	-0.1654	(0.1526)	-0.1943	(0.1793)
PMK female	-0.8001	(0.2860)	-0.8622	(0.7040)
living w/ two parents	-0.4325	(0.4166)	-0.8516	(0.5761)
living w/ biological	0.0900	(0.0775)	0.2713	(0.1732)
hlthc(1)poor			-1.3039	(0.3356)
hlthc(1)fair			-0.6808	(0.1385)
hlthc(1)good			-0.2170	(0.0555)
hlthc(1)excellent			0.5028	(0.0359)
mln(hh income)			0.3091	(0.0534)
magec			0.0006	(0.0096)
mfsize			0.0955	(0.0324)
mschoolm			0.0635	(0.0293)
mschoolf			0.0601	(0.0262)
mpmknm			0.7509	(1.5483)
mpmkfe			0.8813	(1.4943)
mtwopar			1.6287	(2.0270)
mlwbiopa			-0.3432	(0.2486)
msmxmpm			-0.0181	(0.1244)
cut1	-2.5054	(0.5444)	0.6633	(2.4564)
cut2	-1.5663	(0.5372)	1.7581	(2.4554)
cut3	-0.3577	(0.5351)	3.1712	(2.4552)
cut4	0.8338	(0.5351)	4.5678	(2.4554)
ICC			0.3064	(0.0135)
Log likelihood	-16164.3		-15748.5	

1. Standard errors are reported in parentheses. These are robust to cluster effects for the pooled specification.
2. ICC is the intra-class correlation coefficient, $(\sigma_u^2/(1+\sigma_u^2))$

Table 1.6: Average partial effects on probability of reporting excellent health

	(1)	(2)	(3)	(4)
	Pooled model, without correlated effects specifications		Random effects, with correlated effects specifications	
hlthc(t-1)poor	-0.4671	(0.0952)	-0.2712	(0.0340)
hlthc(t-1)fair	-0.2099	(0.0268)	-0.1357	(0.0128)
hlthc(t-1)good	-0.0635	(0.0055)	-0.0723	(0.0065)
hlthc(t-1)excellent	0.0652	(0.0015)	0.0912	(0.0074)
child age	-0.0005	(0.0002)	-0.0009	(0.0001)
child gender	-0.0037	(0.0018)	-0.0148	(0.0015)
family size	0.0026	(0.0009)	-0.0104	(0.0011)
mbirthage	0.0000	(0.0002)	-0.0048	(0.0005)
ln(hh income)	0.0011	(0.0018)	0.0074	(0.0008)
mother school2	0.0114	(0.0029)	0.0309	(0.0034)
mother school3	0.0141	(0.0029)	0.0270	(0.0029)
mother school4	0.0198	(0.0028)	0.0377	(0.0037)
father school2	0.0058	(0.0027)	0.0044	(0.0005)
father school3	0.0050	(0.0028)	-0.0133	(0.0014)
father school4	0.0061	(0.0026)	-0.0246	(0.0027)
PMK not mother	-0.0668	(0.0528)	-0.1488	(0.0161)
mother school2*PMKnm	-0.0274	(0.0210)	-0.0980	(0.0095)
mother school3*PMKnm	-0.0163	(0.0174)	-0.0476	(0.0047)
mother school4*PMKnm	-0.0163	(0.0168)	-0.0590	(0.0059)
PMK female	-0.0448	(0.0274)	-0.2278	(0.0419)
living w/ two parents	-0.0283	(0.0369)	-0.2229	(0.0413)
living w/ biological parents	0.0085	(0.0069)	0.0827	(0.0080)
hlthc(1)poor			-0.3708	(0.0579)
hlthc(1)fair			-0.2075	(0.0207)
hlthc(1)good			-0.0664	(0.0061)
hlthc(1)excellent			0.1575	(0.0103)
mln(hh income)			0.0928	(0.0097)
magec			0.0002	(0.0000)
mfsize			0.0287	(0.0030)
mschoolm			0.0191	(0.0020)
mschoolf			0.0180	(0.0019)
mpmknm			0.2254	(0.0236)
mpmkfe			0.2646	(0.0277)
mtwopar			0.4890	(0.0511)
mlwbiopa			-0.1030	(0.0108)
msmxmpm			-0.0054	(0.0006)

1. Standard errors are reported in parentheses.

Table 1.7: Transition matrix for empirical model, balanced study sample

		Fair/Poor	Good	Very Good	Excellent
		t	t	t	t
Fair/Poor	t-1	0.089	0.289	0.391	0.231
Good	t-1	0.030	0.191	0.409	0.369
Very Good	t-1	0.013	0.121	0.368	0.498
Excellent	t-1	0.004	0.058	0.274	0.664

Table 1.8a: Average Partial effects for the probability of reporting excellent health by neighbourhood status, random effects model-- by quartiles of mean household income

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Part A: APE and standard errors by quartiles of mean household income of neighbourhood								
	Lowest income		Second lowest income		Middle income		Highest income	
hlthc(t-1)poor/fair	-0.0959	(0.0102)	-0.1691	(0.0166)	-0.1528	(0.0167)	-0.1663	(0.0155)
hlthc(t-1)good	-0.0446	(0.0043)	-0.0401	(0.0040)	-0.0961	(0.0102)	-0.1025	(0.0106)
hlthc(t-1)excellent	0.0868	(0.0071)	0.1201	(0.0090)	0.0756	(0.0077)	0.0760	(0.0087)
child age	-0.0011	(0.0001)	0.0007	(0.0001)	-0.0047	(0.0006)	0.0015	(0.0002)
child gender	-0.0003	(0.0000)	-0.0046	(0.0005)	-0.0151	(0.0018)	-0.0322	(0.0045)
family size	-0.0147	(0.0015)	0.0108	(0.0011)	-0.0093	(0.0011)	-0.0272	(0.0038)
mbirthage	-0.0052	(0.0005)	-0.0068	(0.0007)	-0.0010	(0.0001)	-0.0062	(0.0009)
ln(hh income)	0.0000	(0.0000)	-0.0177	(0.0019)	0.0380	(0.0045)	-0.0032	(0.0004)
lwbiopa	0.0984	(0.0108)	0.1241	(0.0127)	-0.0363	(0.0044)	0.1159	(0.0121)
mother school2	0.0165	(0.0017)	0.0388	(0.0044)	0.0477	(0.0060)	0.0124	(0.0018)
mother school3	-0.0244	(0.0025)	0.0321	(0.0036)	0.0922	(0.0120)	-0.0101	(0.0014)
mother school4	0.0021	(0.0002)	0.0725	(0.0072)	0.0744	(0.0088)	-0.0217	(0.0032)
father school2	0.0052	(0.0005)	-0.0021	(0.0002)	-0.0152	(0.0018)	0.0249	(0.0037)
father school3	-0.0068	(0.0007)	-0.0019	(0.0002)	-0.0659	(0.0077)	0.0263	(0.0039)
father school4	-0.0376	(0.0041)	-0.0242	(0.0026)	-0.0723	(0.0091)	0.0346	(0.0047)
PMK not mother	0.0660	(0.0073)	0.1729	(0.0266)	0.1180	(0.0165)	0.2410	(0.0657)
schoolm2*PMKnm	0.0735	(0.0082)	-0.1995	(0.0230)	-0.1537	(0.0180)	-0.4098	(0.0550)
schoolm3*PMKnm	0.1289	(0.0162)	-0.1567	(0.0170)	-0.1422	(0.0169)	-0.2237	(0.0229)
schoolm4*PMKnm	-0.0243	(0.0025)	-0.1444	(0.0159)	-0.0695	(0.0081)	-0.2449	(0.0287)
hlthc(1)poor/fair	-0.0686	(0.0071)	-0.2300	(0.0236)	-0.3884	(0.0564)	-0.2846	(0.0261)
hlthc(1)good	-0.0914	(0.0084)	-0.0810	(0.0077)	-0.0547	(0.0061)	-0.0096	(0.0013)
hlthc(1)excellent	0.1783	(0.0109)	0.1359	(0.0098)	0.1537	(0.0133)	0.1929	(0.0148)
mln(hh income)	0.0558	(0.0058)	0.1109	(0.0118)	0.1057	(0.0124)	0.0962	(0.0136)
mlwbiopa	-0.0726	(0.0075)	-0.1646	(0.0175)	-0.0211	(0.0025)	-0.0559	(0.0079)
magec	0.0012	(0.0001)	0.0093	(0.0010)	-0.0010	(0.0001)	-0.0072	(0.0010)
mfsize	0.0395	(0.0041)	0.0045	(0.0005)	0.0167	(0.0020)	0.0566	(0.0080)
mschoolm	0.0434	(0.0045)	0.0014	(0.0001)	-0.0148	(0.0017)	0.0337	(0.0047)
mschoolf	0.0336	(0.0035)	0.0218	(0.0023)	0.0187	(0.0022)	-0.0027	(0.0004)
mpmknm	-0.3190	(0.0331)	-0.0608	(0.0064)	-0.0768	(0.0090)	0.4859	(0.0685)
msmxmpm	0.0439	(0.0046)	-0.0005	(0.0001)	0.0291	(0.0034)	-0.1498	(0.0211)

Part B: Test of significance of difference in APE for the probability of reporting excellent health across quartiles of mean household income

	Test-statistic	p-value	Test-statistic	p-value	Test-statistic	p-value
	1st quartile vs. 2nd quartile		1st quartile vs. 3rd quartile		1st quartile vs. 4th quartile	
hlthc(t-1)poor/fair	3.768	0.000	2.906	0.004	3.795	0.000
hlthc(t-1)good	-0.771	0.441	4.647	0.000	5.045	0.000
hlthc(t-1)excellent	-2.911	0.004	1.067	0.286	0.961	0.336
			2nd quartile vs. 3rd quartile		2nd quartile vs. 4th quartile	
hlthc(t-1)poor/fair			-0.692	0.489	-0.125	0.900
hlthc(t-1)good			5.118	0.000	5.503	0.000
hlthc(t-1)excellent			3.771	0.000	3.533	0.000
					3rd quartile vs. 4th quartile	
hlthc(t-1)poor/fair					0.589	0.556
hlthc(t-1)good					0.434	0.664
hlthc(t-1)excellent					-0.032	0.975

Table 1.8b: Average Partial effects for the probability of reporting excellent health by neighbourhood status, random effects model-- by quartiles of proportion of population with university degree in neighbourhood

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Part A: APE and standard errors by quartiles of proportion of population with university degree in neighbourhood								
	Lowest % w/ degree		Second lowest %		Second highest %		Highest % w/ degree	
hlthc(t-1)poor/fair	-0.1604	(0.0159)	-0.0679	(0.0065)	-0.2589	(0.0333)	-0.1238	(0.0139)
hlthc(t-1)good	-0.0634	(0.0058)	-0.0455	(0.0043)	-0.0979	(0.0108)	-0.0786	(0.0092)
hlthc(t-1)excellent	0.0824	(0.0067)	0.0873	(0.0072)	0.0914	(0.0095)	0.1053	(0.0114)
child age	-0.0027	(0.0003)	-0.0041	(0.0004)	0.0004	(0.0001)	0.0026	(0.0004)
child gender	0.0129	(0.0013)	-0.0106	(0.0011)	-0.0336	(0.0043)	-0.0217	(0.0031)
family size	-0.0028	(0.0003)	-0.0053	(0.0005)	-0.0350	(0.0046)	0.0078	(0.0011)
mbirthage	-0.0044	(0.0004)	-0.0048	(0.0005)	-0.0034	(0.0004)	-0.0080	(0.0011)
ln(hh income)	-0.0120	(0.0012)	0.0133	(0.0014)	0.0251	(0.0033)	-0.0084	(0.0012)
lwbiopa	0.0986	(0.0100)	-0.0207	(0.0021)	0.1320	(0.0162)	0.1288	(0.0147)
mother school2	0.0203	(0.0021)	0.0539	(0.0060)	-0.0105	(0.0013)	0.0133	(0.0019)
mother school3	0.0261	(0.0027)	0.0139	(0.0014)	0.0215	(0.0029)	-0.0046	(0.0006)
mother school4	0.0116	(0.0012)	0.0877	(0.0077)	0.0015	(0.0002)	0.0004	(0.0001)
father school2	0.0265	(0.0028)	0.0078	(0.0008)	-0.0448	(0.0055)	0.0361	(0.0056)
father school3	0.0327	(0.0034)	-0.0302	(0.0030)	-0.0585	(0.0075)	0.0162	(0.0024)
father school4	0.0096	(0.0010)	-0.0400	(0.0043)	-0.0807	(0.0118)	0.0024	(0.0003)
PMK not mother	0.0937	(0.0114)	0.1481	(0.0186)	0.0562	(0.0080)	0.2771	(0.0806)
schoolm2*PMKnm	0.0293	(0.0031)	-0.2124	(0.0246)	-0.0181	(0.0023)	-0.4087	(0.0634)
schoolm3*PMKnm	-0.0741	(0.0073)	0.0228	(0.0024)	-0.1031	(0.0125)	-0.2557	(0.0312)
schoolm4*PMKnm	-0.0770	(0.0077)	-0.0490	(0.0050)	0.0031	(0.0004)	-0.3333	(0.0482)
hlthc(1)poor/fair	-0.2033	(0.0210)	-0.2106	(0.0215)	-0.2641	(0.0342)	-0.3071	(0.0350)
hlthc(1)good	-0.0501	(0.0047)	-0.0940	(0.0086)	-0.0402	(0.0049)	-0.0672	(0.0082)
hlthc(1)excellent	0.1839	(0.0106)	0.1112	(0.0084)	0.1685	(0.0141)	0.1695	(0.0156)
mln(hh income)	0.0713	(0.0073)	0.0804	(0.0082)	0.0969	(0.0126)	0.1590	(0.0228)
mlwbiopa	-0.1394	(0.0142)	0.0531	(0.0054)	-0.2257	(0.0294)	-0.1117	(0.0160)
magec	0.0102	(0.0010)	0.0061	(0.0006)	-0.0020	(0.0003)	-0.0142	(0.0020)
mfsize	0.0081	(0.0008)	0.0283	(0.0029)	0.0613	(0.0080)	0.0199	(0.0028)
mschoolm	0.0388	(0.0040)	0.0083	(0.0008)	0.0112	(0.0015)	0.0016	(0.0002)
mschoolf	-0.0006	(0.0001)	0.0261	(0.0027)	0.0385	(0.0050)	0.0155	(0.0022)
mpmknm	-0.1856	(0.0189)	-0.3204	(0.0325)	0.2192	(0.0286)	0.2450	(0.0351)
msmxmpm	0.0151	(0.0015)	0.0920	(0.0093)	-0.0639	(0.0083)	-0.0886	(0.0127)

Part B: Test of significance of difference in APE for prob. of excellent health across quartiles of proportion with university degree						
	Test-statistic	p-value	Test-statistic	p-value	Test-statistic	p-value
	1st quartile vs. 2nd quartile		1st quartile vs. 3rd quartile		1st quartile vs. 4th quartile	
hlthc(t-1)poor/fair	-5.377	0.000	2.668	0.008	-1.734	0.083
hlthc(t-1)good	-2.485	0.013	2.818	0.005	1.388	0.165
hlthc(t-1)excellent	-0.499	0.618	-0.780	0.435	-1.737	0.082
			2ndquartile vs. 3rd quartile		2nd quartile vs. 4th quartile	
hlthc(t-1)poor/fair			5.628	0.000	3.642	0.000
hlthc(t-1)good			4.523	0.000	3.250	0.001
hlthc(t-1)excellent			-0.348	0.728	-1.338	0.181
					3rd quartile vs. 4th quartile	
hlthc(t-1)poor/fair					-3.745	0.000
hlthc(t-1)good					-1.364	0.172
hlthc(t-1)excellent					-0.937	0.349

Table 1.8c: Average Partial effects for the probability of reporting excellent health by neighbourhood status, random effects model-- by quartiles of proportion of families headed by lone-parents in neighborhood

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Part A: APE and standard errors by quartiles of proportion of families headed by lone-parents in neighborhood								
	Highest %		Second highest %		Second lowest %		Lowest %	
hlthc(t-1)poor/fair	-0.1492	(0.0160)	-0.0537	(0.0052)	-0.2530	(0.0286)	-0.1464	(0.0161)
hlthc(t-1)good	-0.0577	(0.0052)	-0.0645	(0.0061)	-0.1151	(0.0130)	-0.0503	(0.0058)
hlthc(t-1)excellent	0.0964	(0.0072)	0.0696	(0.0063)	0.0693	(0.0083)	0.1263	(0.0115)
child age	-0.0004	(0.0000)	-0.0018	(0.0002)	-0.0026	(0.0004)	0.0007	(0.0001)
child gender	-0.0290	(0.0027)	0.0053	(0.0006)	-0.0140	(0.0020)	-0.0067	(0.0009)
family size	0.0019	(0.0002)	-0.0058	(0.0006)	-0.0088	(0.0013)	-0.0249	(0.0032)
mbirthage	-0.0061	(0.0006)	-0.0021	(0.0002)	-0.0047	(0.0007)	-0.0071	(0.0009)
ln(hh income)	0.0042	(0.0004)	0.0291	(0.0031)	-0.0087	(0.0012)	-0.0008	(0.0001)
lwbiopa	0.0448	(0.0044)	0.0677	(0.0067)	0.2354	(0.0295)	-0.2998	(0.0860)
mother school2	0.0322	(0.0031)	0.0562	(0.0067)	0.0422	(0.0066)	-0.0086	(0.0011)
mother school3	0.0212	(0.0020)	0.0532	(0.0062)	0.0536	(0.0084)	-0.0309	(0.0039)
mother school4	0.0559	(0.0050)	0.0513	(0.0052)	0.0681	(0.0093)	-0.0382	(0.0053)
father school2	0.0005	(0.0001)	0.0538	(0.0061)	-0.0153	(0.0021)	-0.0229	(0.0029)
father school3	-0.0234	(0.0022)	0.0197	(0.0022)	-0.0204	(0.0029)	-0.0247	(0.0031)
father school4	-0.0422	(0.0042)	0.0168	(0.0018)	-0.0295	(0.0044)	-0.0486	(0.0068)
PMK not mother	0.0613	(0.0058)	0.2071	(0.0360)	0.1462	(0.0278)	0.0418	(0.0058)
schoolm2*PMKnm	-0.0902	(0.0094)	-0.1818	(0.0185)	-0.2226	(0.0273)	0.0722	(0.0110)
schoolm3*PMKnm	0.1015	(0.0101)	-0.2747	(0.0336)	0.0267	(0.0040)	-0.0122	(0.0015)
schoolm4*PMKnm	-0.0309	(0.0030)	-0.1888	(0.0213)	-0.0848	(0.0112)	0.0497	(0.0071)
hlthc(1)poor/fair	-0.2561	(0.0335)	-0.2882	(0.0312)	-0.2370	(0.0252)	-0.1693	(0.0188)
hlthc(1)good	-0.1209	(0.0108)	-0.0014	(0.0002)	-0.0933	(0.0112)	-0.0172	(0.0021)
hlthc(1)excellent	0.1205	(0.0084)	0.1790	(0.0116)	0.1498	(0.0141)	0.1999	(0.0141)
mln(hh income)	0.0952	(0.0090)	0.0926	(0.0099)	0.0961	(0.0138)	0.0846	(0.0110)
mlwbiopa	-0.0985	(0.0093)	-0.0065	(0.0007)	-0.2204	(0.0317)	0.2425	(0.0314)
magec	-0.0066	(0.0006)	0.0074	(0.0008)	0.0037	(0.0005)	-0.0024	(0.0003)
mfsize	0.0263	(0.0025)	0.0151	(0.0016)	0.0276	(0.0040)	0.0444	(0.0057)
mschoolm	0.0045	(0.0004)	0.0101	(0.0011)	0.0022	(0.0003)	0.0486	(0.0063)
mschoolf	0.0299	(0.0028)	-0.0056	(0.0006)	0.0271	(0.0039)	0.0308	(0.0040)
mpmknm	0.1636	(0.0155)	-0.3126	(0.0332)	-0.0622	(0.0089)	-0.1976	(0.0256)
msmxmpm	-0.0743	(0.0070)	0.0683	(0.0073)	0.0276	(0.0040)	0.0291	(0.0038)
Part B: Test of significance of difference in APE for prob. of excellent health across quartiles of proportion of lone-parent families								
			Test-statistic	p-value	Test-statistic	p-value	Test-statistic	p-value
			1st quartile vs. 2nd quartile		1st quartile vs. 3rd quartile		1st quartile vs. 4th quartile	
hlthc(t-1)poor/fair			-5.661	0.000	3.168	0.002	-0.122	0.903
hlthc(t-1)good			0.844	0.399	4.112	0.000	-0.956	0.339
hlthc(t-1)excellent			2.812	0.005	2.461	0.014	-2.206	0.027
					2ndquartile vs. 3rd quartile		2nd quartile vs. 4th quartile	
hlthc(t-1)poor/fair					6.864	0.000	5.494	0.000
hlthc(t-1)good					3.542	0.000	-1.688	0.091
hlthc(t-1)excellent					0.028	0.978	-4.325	0.000
							3rd quartile vs. 4th quartile	
hlthc(t-1)poor/fair							-3.252	0.001
hlthc(t-1)good							-4.565	0.000
hlthc(t-1)excellent							-4.008	0.000

Table 1.8d: Average Partial effects for the probability of reporting excellent health by neighbourhood status, random effects model-- by quartiles of proportion of families living in rental accommodations in neighborhood

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Part A: APE and standard errors by quartiles of proportion of families living in rental accommodations in neighborhood								
	Highest % with rental		Second highest %		Second lowest %		Lowest % with rental	
hlthc(t-1)poor/fair	-0.0928	(0.0080)	-0.1679	(0.0184)	-0.1819	(0.0220)	-0.1588	(0.0157)
hlthc(t-1)good	-0.0529	(0.0044)	-0.0982	(0.0100)	-0.0626	(0.0081)	-0.0626	(0.0062)
hlthc(t-1)excellent	0.0938	(0.0066)	0.0889	(0.0084)	0.0917	(0.0108)	0.0755	(0.0070)
child age	0.0015	(0.0001)	-0.0006	(0.0001)	-0.0055	(0.0008)	0.0009	(0.0001)
child gender	-0.0108	(0.0010)	-0.0093	(0.0011)	-0.0150	(0.0022)	-0.0160	(0.0018)
family size	0.0127	(0.0012)	0.0018	(0.0002)	-0.0348	(0.0051)	-0.0210	(0.0024)
mbirthage	-0.0052	(0.0005)	-0.0004	(0.0000)	-0.0119	(0.0018)	-0.0029	(0.0003)
ln(hh income)	-0.0102	(0.0009)	0.0062	(0.0007)	0.0190	(0.0028)	0.0015	(0.0002)
lwbiopa	0.0416	(0.0037)	0.1000	(0.0114)	0.0715	(0.0096)	-0.1688	(0.0305)
mother school2	0.0474	(0.0047)	0.0512	(0.0067)	0.0379	(0.0060)	-0.0141	(0.0016)
mother school3	0.0458	(0.0045)	0.0652	(0.0085)	0.0377	(0.0059)	-0.0453	(0.0049)
mother school4	0.1020	(0.0088)	0.0501	(0.0057)	0.0501	(0.0072)	-0.0595	(0.0078)
father school2	0.0246	(0.0023)	0.0123	(0.0015)	-0.0274	(0.0039)	-0.0046	(0.0005)
father school3	0.0046	(0.0004)	-0.0126	(0.0015)	-0.0703	(0.0099)	0.0085	(0.0010)
father school4	-0.0183	(0.0017)	0.0041	(0.0005)	-0.1001	(0.0170)	-0.0115	(0.0013)
PMK not mother	0.1665	(0.0216)	0.2067	(0.0367)	0.0022	(0.0003)	0.0518	(0.0065)
schoolm2*PMKnm	-0.1561	(0.0152)	-0.2032	(0.0254)	-0.0039	(0.0006)	-0.0685	(0.0070)
schoolm3*PMKnm	-0.0425	(0.0038)	-0.3142	(0.0475)	0.1783	(0.0394)	-0.0470	(0.0050)
schoolm4*PMKnm	-0.1330	(0.0130)	-0.2019	(0.0268)	0.1096	(0.0204)	-0.0236	(0.0026)
hlthc(1)poor/fair	-0.2950	(0.0339)	-0.2549	(0.0295)	-0.2752	(0.0362)	-0.1293	(0.0122)
hlthc(1)good	-0.0923	(0.0075)	-0.1039	(0.0107)	-0.0728	(0.0094)	0.0154	(0.0018)
hlthc(1)excellent	0.1289	(0.0082)	0.1130	(0.0100)	0.1745	(0.0162)	0.2264	(0.0125)
mln(hh income)	0.1165	(0.0106)	0.1282	(0.0154)	0.0995	(0.0147)	0.0707	(0.0080)
mlwbiopa	-0.0825	(0.0075)	-0.0980	(0.0117)	-0.0827	(0.0122)	0.0541	(0.0062)
magec	-0.0102	(0.0009)	0.0064	(0.0008)	0.0033	(0.0005)	0.0002	(0.0000)
mfsize	0.0013	(0.0001)	0.0122	(0.0015)	0.0741	(0.0110)	0.0296	(0.0034)
mschoolm	-0.0209	(0.0019)	0.0097	(0.0012)	0.0083	(0.0012)	0.0617	(0.0070)
mschoolf	0.0252	(0.0023)	0.0040	(0.0005)	0.0519	(0.0077)	0.0178	(0.0020)
mpmknm	-0.1853	(0.0169)	-0.2155	(0.0258)	0.1825	(0.0270)	-0.0139	(0.0016)
msmxmpm	0.0272	(0.0025)	0.0681	(0.0082)	-0.1249	(0.0185)	0.0139	(0.0016)
Part B: Test of significance of difference in APE for prob. of excellent health across quartiles of % of families living in rental accommodations								
			Test-statistic	p-value	Test-statistic	p-value	Test-statistic	p-value
			1st quartile vs. 2nd quartile		1st quartile vs. 3rd quartile		1st quartile vs. 4th quartile	
hlthc(t-1)poor/fair			3.752	0.000	3.804	0.000	3.756	0.000
hlthc(t-1)good			4.155	0.000	1.049	0.294	1.276	0.202
hlthc(t-1)excellent			0.460	0.646	0.168	0.866	1.903	0.057
					2ndquartile vs. 3rd quartile		2nd quartile vs. 4th quartile	
hlthc(t-1)poor/fair					0.487	0.626	-0.376	0.707
hlthc(t-1)good					-2.762	0.006	-3.033	0.002
hlthc(t-1)excellent					-0.204	0.838	1.221	0.222
							3rd quartile vs. 4th quartile	
hlthc(t-1)poor/fair							-0.852	0.394
hlthc(t-1)good							-0.003	0.998
hlthc(t-1)excellent							1.258	0.208

Table 1.9: Transition matrices by long-term neighbourhood status for empirical model

By quartiles of mean household income of neighbourhood															
Lowest income				Second lowest income				Middle income				Highest income			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.251	0.416	0.333	<=Good	0.226	0.420	0.354	<=Good	0.272	0.389	0.340	<=Good	0.207	0.401	0.392
Very good	0.153	0.386	0.462	Very good	0.139	0.390	0.471	Very good	0.133	0.355	0.512	Very good	0.110	0.340	0.549
Excellent	0.075	0.300	0.625	Excellent	0.058	0.281	0.660	Excellent	0.067	0.275	0.657	Excellent	0.046	0.237	0.717
By quartiles of proportion of population with university degree in neighbourhood															
Lowest % with college degree				Second lowest %				Second highest %				Highest % with college degree			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.252	0.409	0.339	<=Good	0.246	0.387	0.367	<=Good	0.255	0.423	0.322	<=Good	0.207	0.420	0.372
Very good	0.143	0.373	0.484	Very good	0.156	0.359	0.485	Very good	0.128	0.380	0.492	Very good	0.112	0.369	0.519
Excellent	0.069	0.283	0.648	Excellent	0.083	0.286	0.631	Excellent	0.053	0.271	0.676	Excellent	0.042	0.249	0.709
By quartiles of proportion of families headed by lone-parents in neighborhood															
Highest % with lone-parents				Second highest %				Second lowest %				Lowest % with lone-parents			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.268	0.398	0.333	<=Good	0.215	0.400	0.385	<=Good	0.254	0.411	0.336	<=Good	0.227	0.427	0.346
Very good	0.156	0.374	0.470	Very good	0.132	0.365	0.503	Very good	0.109	0.347	0.543	Very good	0.140	0.393	0.467
Excellent	0.082	0.298	0.620	Excellent	0.066	0.282	0.652	Excellent	0.050	0.250	0.700	Excellent	0.048	0.262	0.690
By quartiles of proportion of families living in rental accommodations in neighborhood															
Highest % with rental accommodations				Second highest %				Second lowest %				Lowest % with rental accommodations			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.233	0.393	0.374	<=Good	0.265	0.406	0.329	<=Good	0.248	0.423	0.329	<=Good	0.221	0.413	0.367
Very good	0.146	0.365	0.490	Very good	0.129	0.362	0.508	Very good	0.127	0.375	0.498	Very good	0.138	0.376	0.486
Excellent	0.074	0.284	0.642	Excellent	0.063	0.275	0.662	Excellent	0.051	0.261	0.688	Excellent	0.057	0.270	0.673

Table 1.10: Transition matrices by neighbourhood status for empirical model—stayers across six cycles

By quartiles of mean household income of neighbourhood															
Lowest income				Second lowest income				Middle income				Highest income			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.252	0.433	0.315	<=Good	0.248	0.436	0.316	<=Good	0.289	0.386	0.325	<=Good	0.203	0.412	0.385
Very good	0.152	0.403	0.445	Very good	0.157	0.411	0.432	Very good	0.110	0.343	0.546	Very good	0.115	0.360	0.525
Excellent	0.074	0.316	0.610	Excellent	0.058	0.286	0.656	Excellent	0.067	0.285	0.648	Excellent	0.039	0.230	0.731
By quartiles of proportion of population with university degree in neighbourhood															
Lowest % with college degree				Second lowest %				Second highest %				Highest % with college degree			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.271	0.410	0.319	<=Good	0.264	0.390	0.347	<=Good	0.229	0.426	0.344	<=Good	0.194	0.415	0.391
Very good	0.149	0.376	0.474	Very good	0.157	0.355	0.488	Very good	0.146	0.398	0.457	Very good	0.118	0.386	0.496
Excellent	0.068	0.281	0.651	Excellent	0.084	0.287	0.629	Excellent	0.053	0.276	0.671	Excellent	0.037	0.244	0.719
By quartiles of proportion of families headed by lone-parents in neighborhood															
Highest % with lone-parents				Second highest %				Second lowest %				Lowest % with lone-parents			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.301	0.402	0.297	<=Good	0.215	0.377	0.408	<=Good	0.255	0.392	0.352	<=Good	0.245	0.442	0.313
Very good	0.154	0.382	0.465	Very good	0.132	0.348	0.520	Very good	0.119	0.373	0.508	Very good	0.142	0.397	0.461
Excellent	0.086	0.312	0.601	Excellent	0.057	0.255	0.688	Excellent	0.036	0.226	0.738	Excellent	0.044	0.257	0.699
By quartiles of proportion of families living in rental accommodations in neighborhood															
Highest % with rental accommodations				Second highest %				Second lowest %				Lowest % with rental accommodations			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.241	0.391	0.369	<=Good	0.268	0.379	0.353	<=Good	0.287	0.429	0.284	<=Good	0.251	0.423	0.326
Very good	0.148	0.368	0.484	Very good	0.107	0.317	0.575	Very good	0.146	0.390	0.464	Very good	0.144	0.397	0.459
Excellent	0.069	0.277	0.654	Excellent	0.065	0.260	0.675	Excellent	0.047	0.249	0.704	Excellent	0.048	0.267	0.685

Table 1.11: Transition matrices by neighbourhood transition patterns for empirical model—movers across six cycles

By transition patterns of neighbourhood mean household income

No change				Sliding-down pattern				Climbing-up pattern				Bouncing pattern			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.235	0.417	0.347	<=Good	0.231	0.418	0.352	<=Good	0.254	0.394	0.352	<=Good	0.252	0.387	0.362
Very good	0.133	0.377	0.490	Very good	0.131	0.380	0.489	Very good	0.147	0.360	0.492	Very good	0.132	0.346	0.522
Excellent	0.060	0.279	0.661	Excellent	0.052	0.267	0.681	Excellent	0.064	0.261	0.676	Excellent	0.072	0.277	0.651

By transition patterns of neighbourhood education (proportion of population with university degree in neighbourhood)

No change				Sliding-down pattern				Climbing-up pattern				Bouncing pattern			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.231	0.408	0.361	<=Good	0.269	0.397	0.334	<=Good	0.229	0.407	0.365	<=Good	0.240	0.412	0.348
Very good	0.137	0.376	0.487	Very good	0.142	0.360	0.497	Very good	0.131	0.358	0.511	Very good	0.126	0.375	0.500
Excellent	0.058	0.271	0.670	Excellent	0.064	0.267	0.669	Excellent	0.063	0.268	0.669	Excellent	0.062	0.288	0.650

By transition patterns of neighbourhood lone-parents status (proportion of families headed by lone-parents in neighborhood)

No change				Sliding-down pattern				Climbing-up pattern				Bouncing pattern			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.251	0.401	0.348	<=Good	0.232	0.419	0.349	<=Good	0.255	0.409	0.336	<=Good	0.232	0.400	0.368
Very good	0.136	0.370	0.494	Very good	0.134	0.380	0.486	Very good	0.138	0.363	0.499	Very good	0.131	0.361	0.509
Excellent	0.061	0.273	0.667	Excellent	0.059	0.278	0.663	Excellent	0.057	0.256	0.687	Excellent	0.068	0.282	0.650

By transition patterns of neighbourhood living arrangements (proportion of families living in rental accommodations in neighborhood)

No change				Sliding-down pattern				Climbing-up pattern				Bouncing pattern			
	<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex		<=Good	Very good	Ex
<=Good	0.241	0.401	0.357	<=Good	0.231	0.411	0.358	<=Good	0.255	0.431	0.314	<=Good	0.240	0.388	0.372
Very good	0.135	0.367	0.498	Very good	0.126	0.370	0.504	Very good	0.141	0.391	0.467	Very good	0.139	0.350	0.512
Excellent	0.061	0.271	0.668	Excellent	0.053	0.268	0.678	Excellent	0.060	0.283	0.657	Excellent	0.071	0.271	0.657

1.A1 Appendix: Variable names and definitions

Table 1.A1 Variable names and definitions

Variable Name	Definition
hlthc	Health status of child, 5 categories: excellent, very good, good, fair and poor; hlthc(t-1) refers to the health status in the previous period, e.g. hlthc(t-1)poor is a dummy indicating reported poor health in the last wave; hlthc(1) refers to the reported health status in the initial period, e.g. hlthc(1)good is a dummy indicating reported good health in the first wave.
child age	Age of child
child gender	Gender of child(Male=1)
family size	Total number of persons living in the household
mbirthage	Age of mother at birth of the child
hh income	Total household income from all sources in the past 12 months; ln(hh income) is the log the household income.
mother school	Female caregiver education, 1= less than secondary, 2=secondary school graduation, 3=some post-secondary, 4=college or university degree
father school	Male caregiver education, 1= less than secondary, 2=secondary school graduation, 3=some post-secondary, 4=college or university degree
PMK not mother	Dummy indicating PMK is not the biological mother of the child
mother school*PMKnm	Interaction terms between mother education status (4 categories) and the dummy indicating PMK not the biological mother of child
PMK female	Dummy indicating PMK is female
living w/ two parents	Dummy indicating child living with both parents (including biological parents or any other parental figures in the household)
living w/ biological parents	Dummy indicating child living with both biological parents
area	Province of residence
mln(hh income)	Mean of the log household income variable over the sample period—within individual mean of the ln(hh income) term
magec	Mean of the child age variable over the sample period—within individual mean of child age
mfize	Mean of the family size variable over the sample period—within individual mean of family size
mschoolm	Mean of the female caregiver education variable over the sample period—within individual mean of mother's education status
mschoolf	Mean of the male caregiver education variable over the sample period—within individual mean of father's education status
mpmknm	Mean of the PMK not biological mother dummy variable over the sample period—within individual mean of the "PMK not mother" term
mpmkfe	Mean of the female PMK dummy variable over the sample period—within individual mean of the "PMK female" term
mtwopar	Mean of the living with two parents dummy variable over the sample period—within individual mean of the "living w/ two parents" term
mlwbiopa	Mean of the dummy variable indicating child living with both biological parents over the sample period—within individual mean of the "living w/ biological parents" term
msmxmpm	Mean of the interactions term between mother's education and the dummy indicating PMK not the biological mother over the sample period—within individual mean of the "schoolm*PMKnm" term

1.A2 Appendix: Descriptive Figures

Figure 1.A1: Health status by cycle

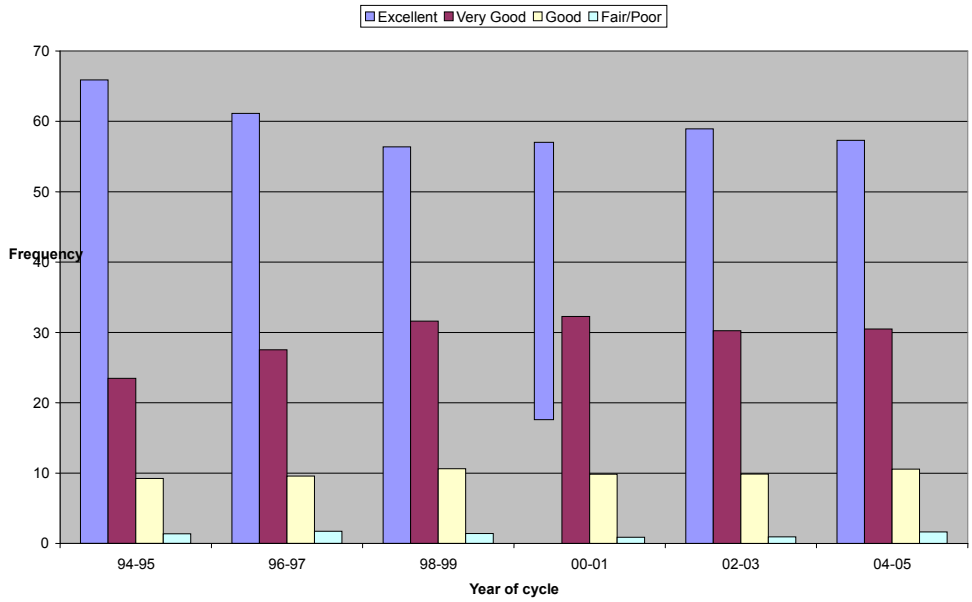


Figure 1.A2: Health status by income class

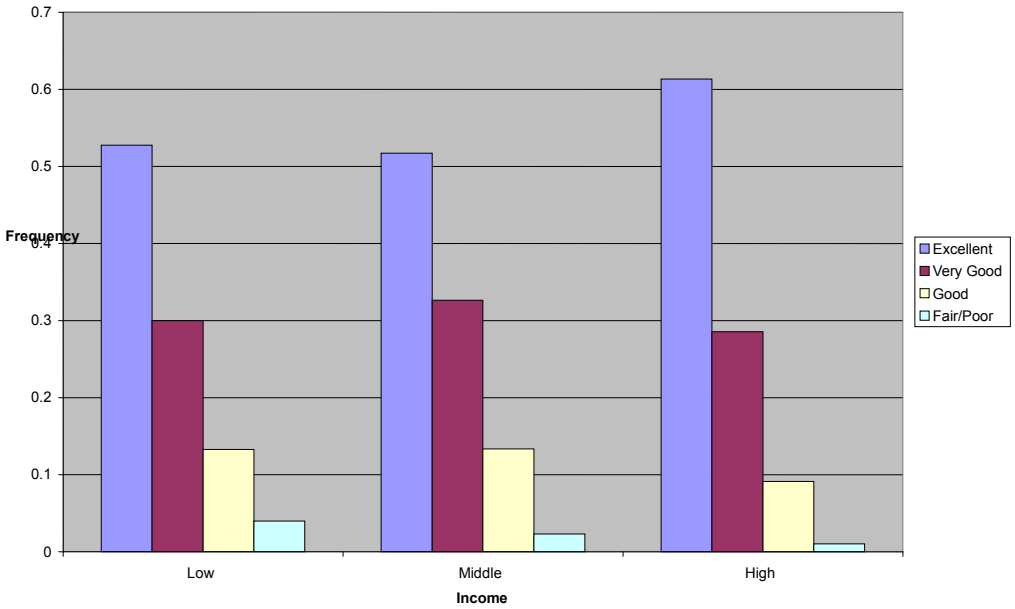


Figure 1.A3: Health status by mother's education

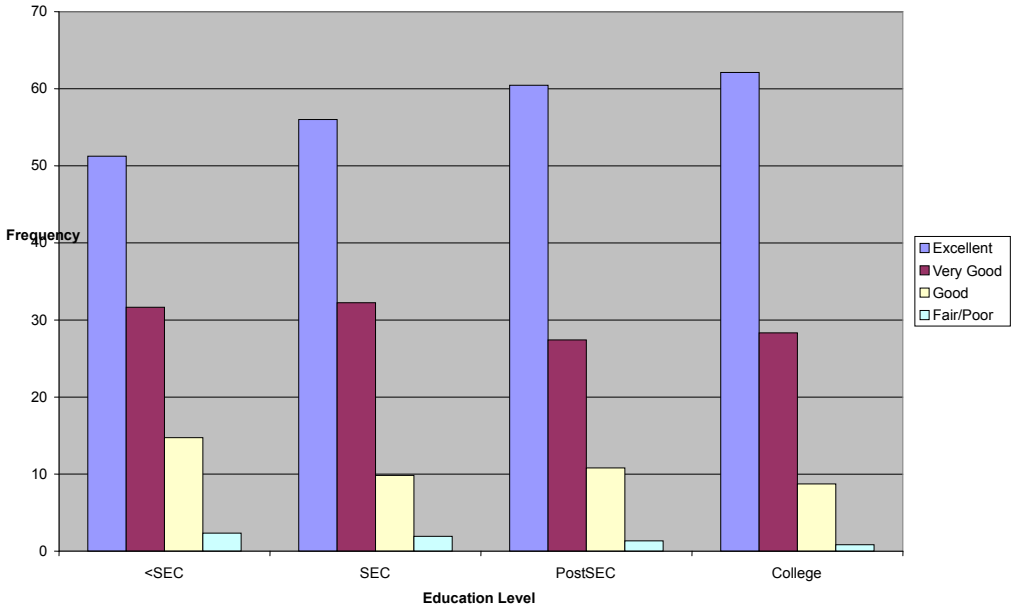


Figure 1.A4: Health Status at cycle 2 by health status at cycle 1

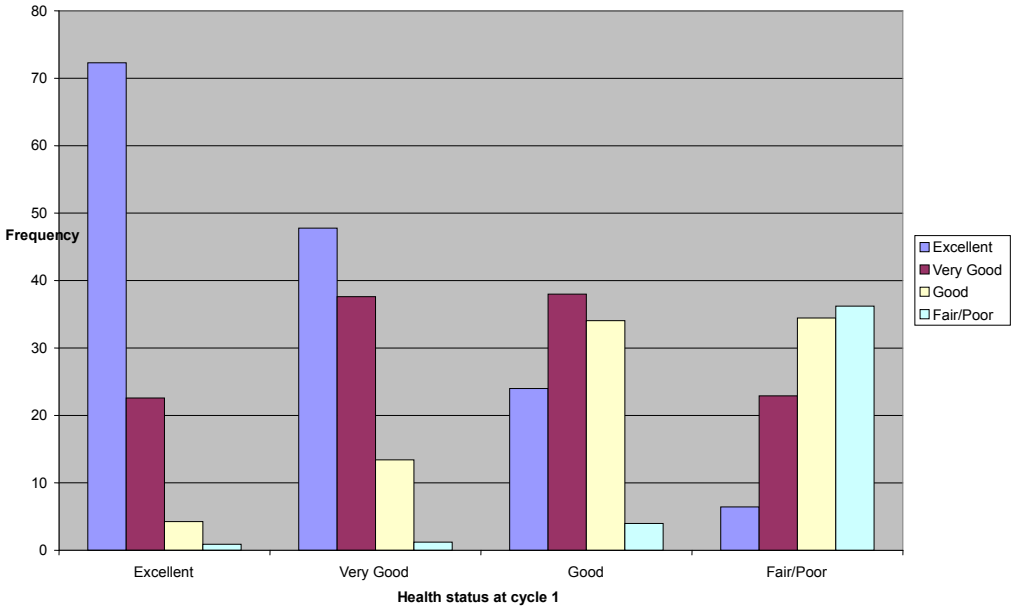
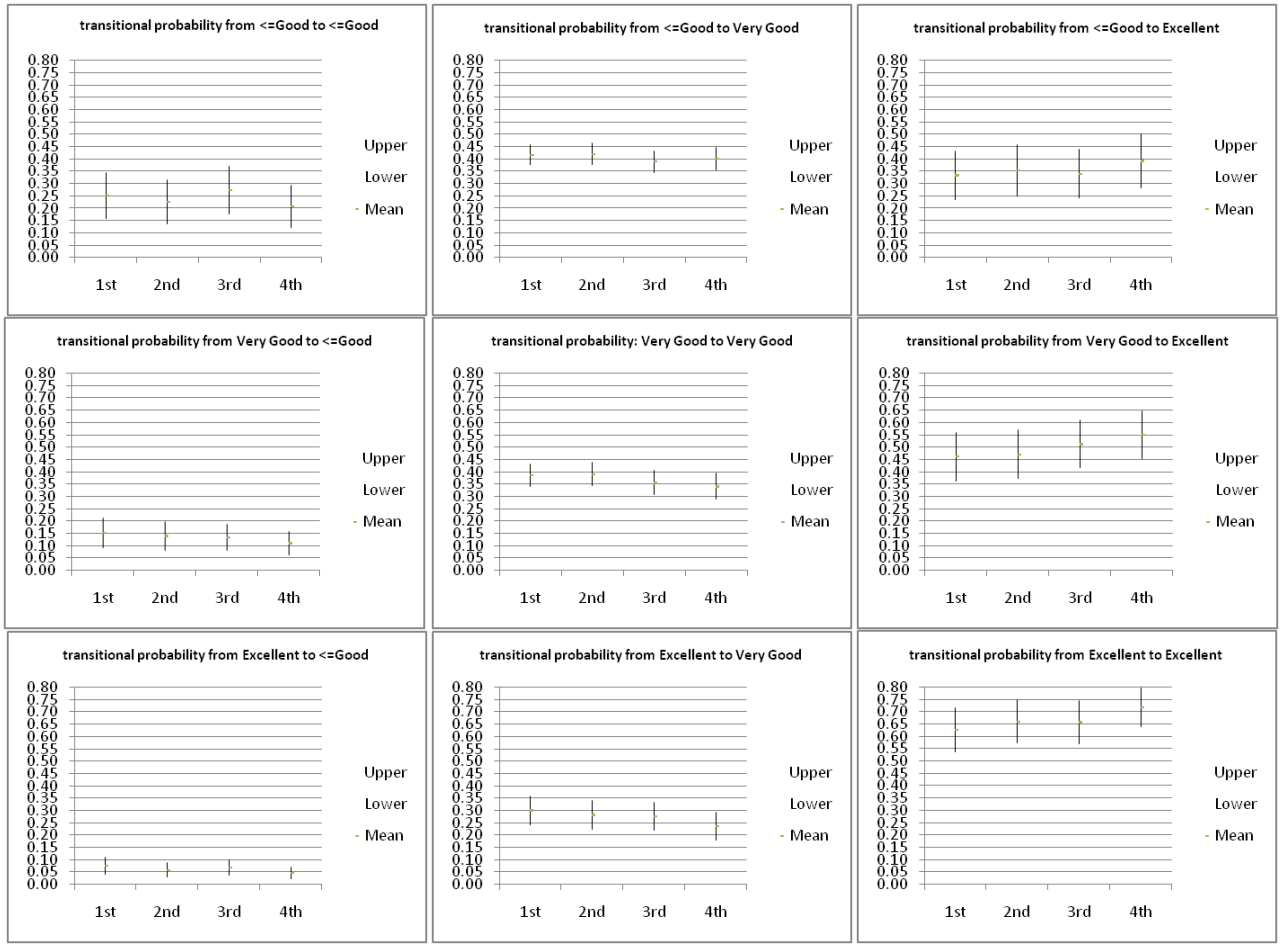
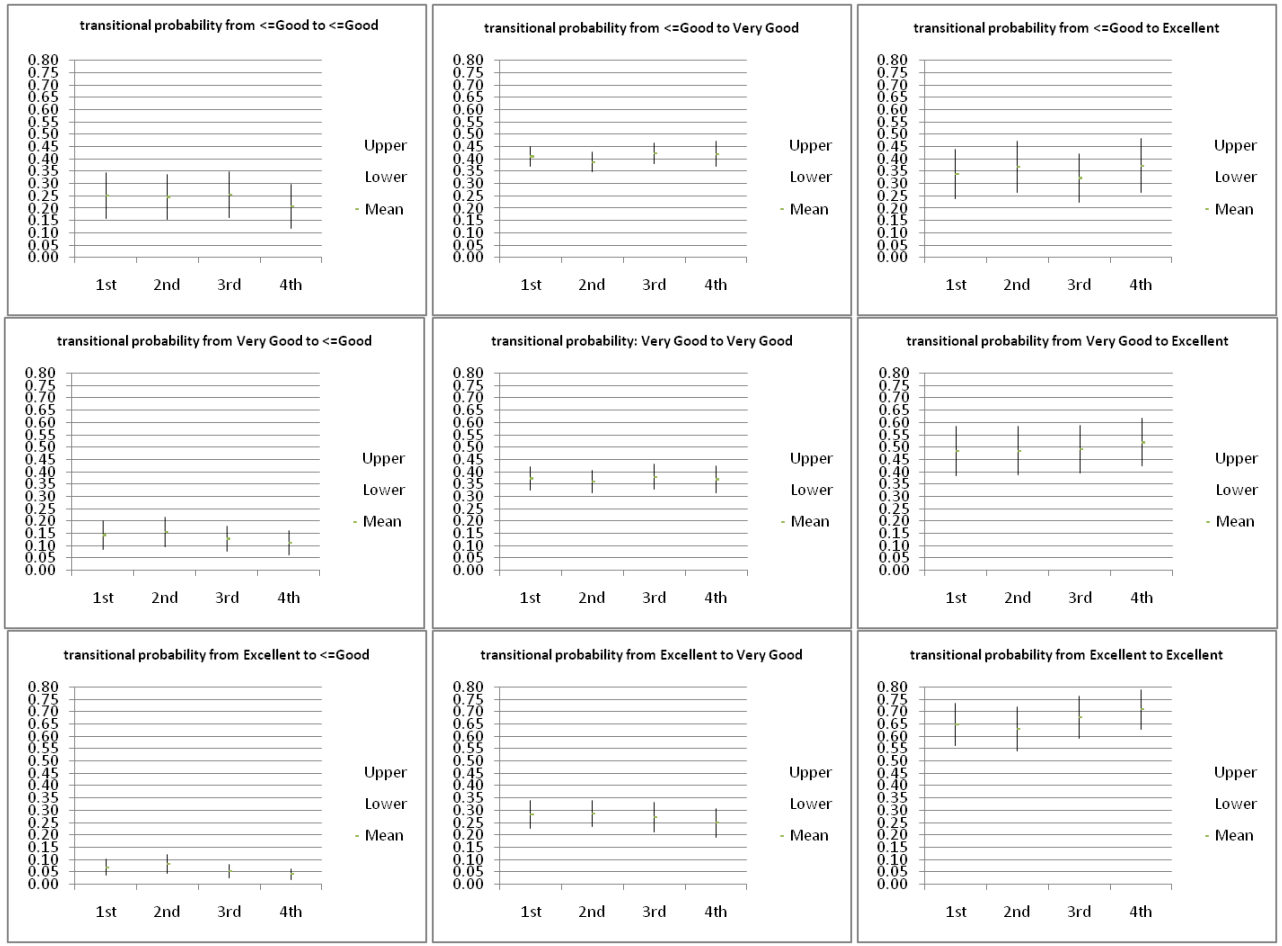


Figure 1.A5 Predicted transitional probabilities based on random effects model by quartiles of neighborhood mean household income



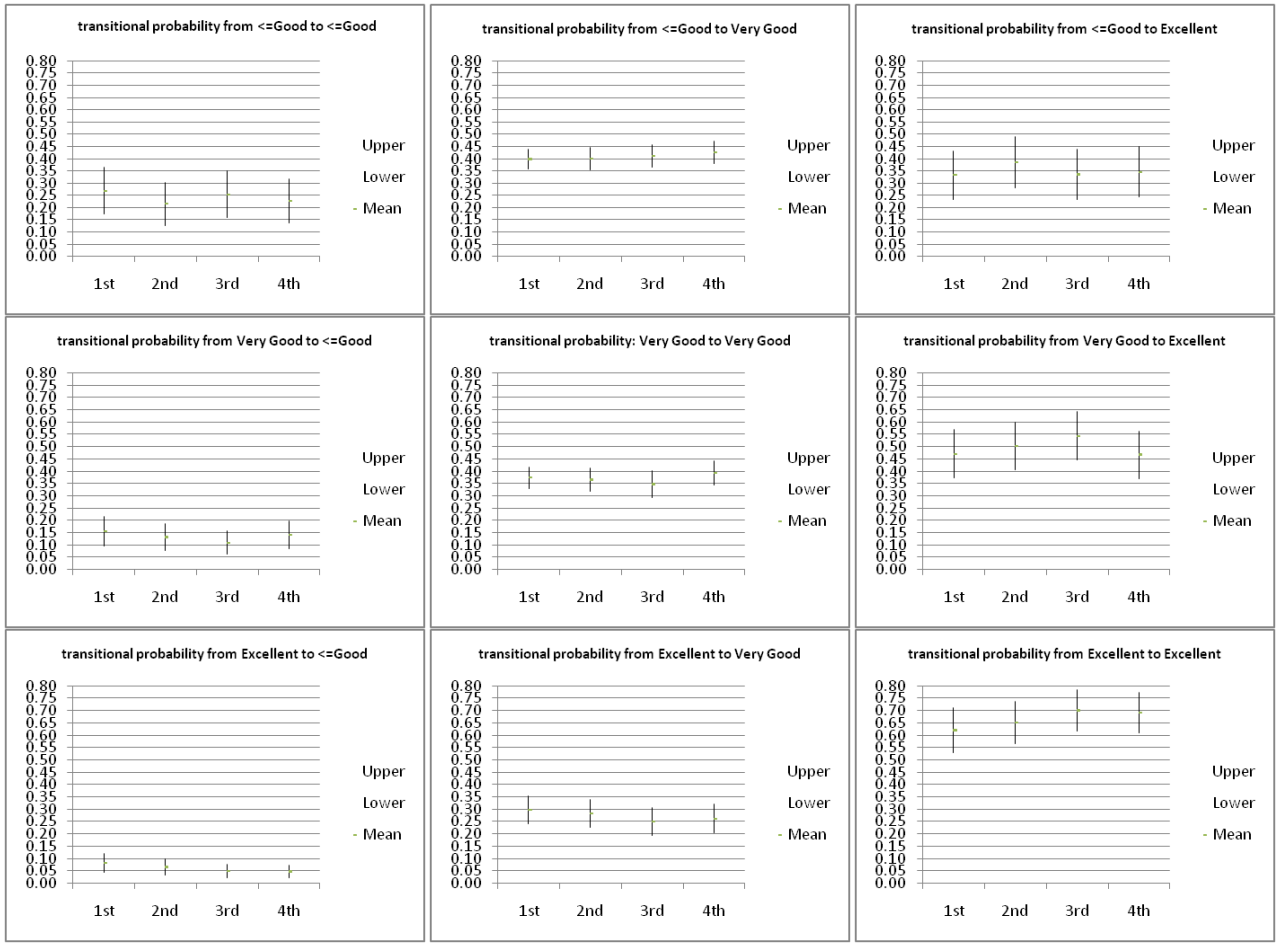
Note: horizontal axis presents lowest income, second lowest, middle income and highest income neighbourhoods, respectively. Mean represents the point estimate of the transitional probability, the vertical line represents the 95% confidence interval of the point estimate.

Figure 1.A6 Predicted transitional probabilities based on random effects model by quartiles of neighborhood education



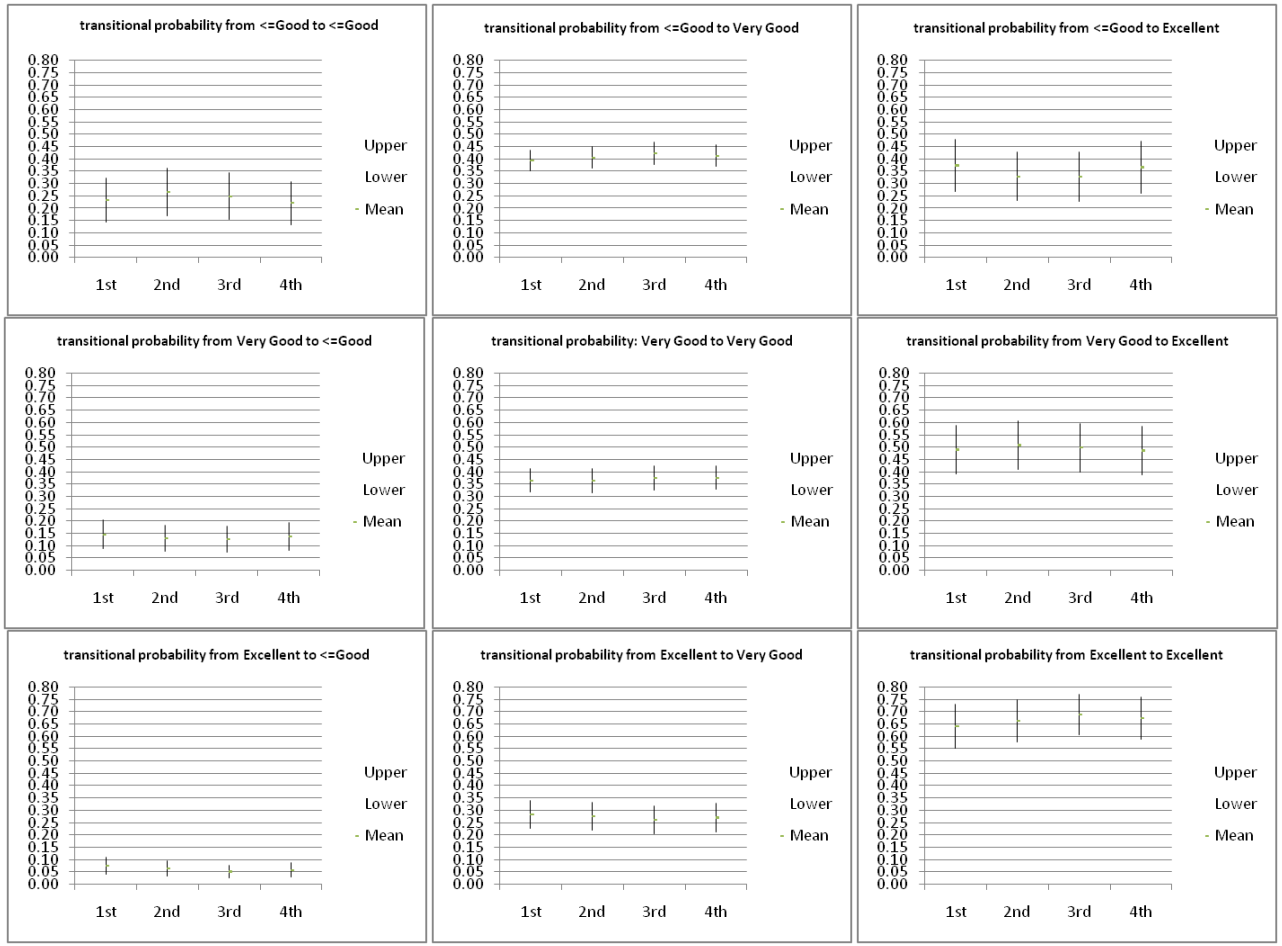
Note: horizontal axis presents lowest %, second lowest %, second highest % and highest % with college degree in neighbourhoods, respectively. Mean represents the point estimate of the transitional probability, the vertical line represents the 95% confidence interval of the point estimate.

Figure 1.A7 Predicted transitional probabilities based on random effects model by quartiles of neighborhood lone-parents status



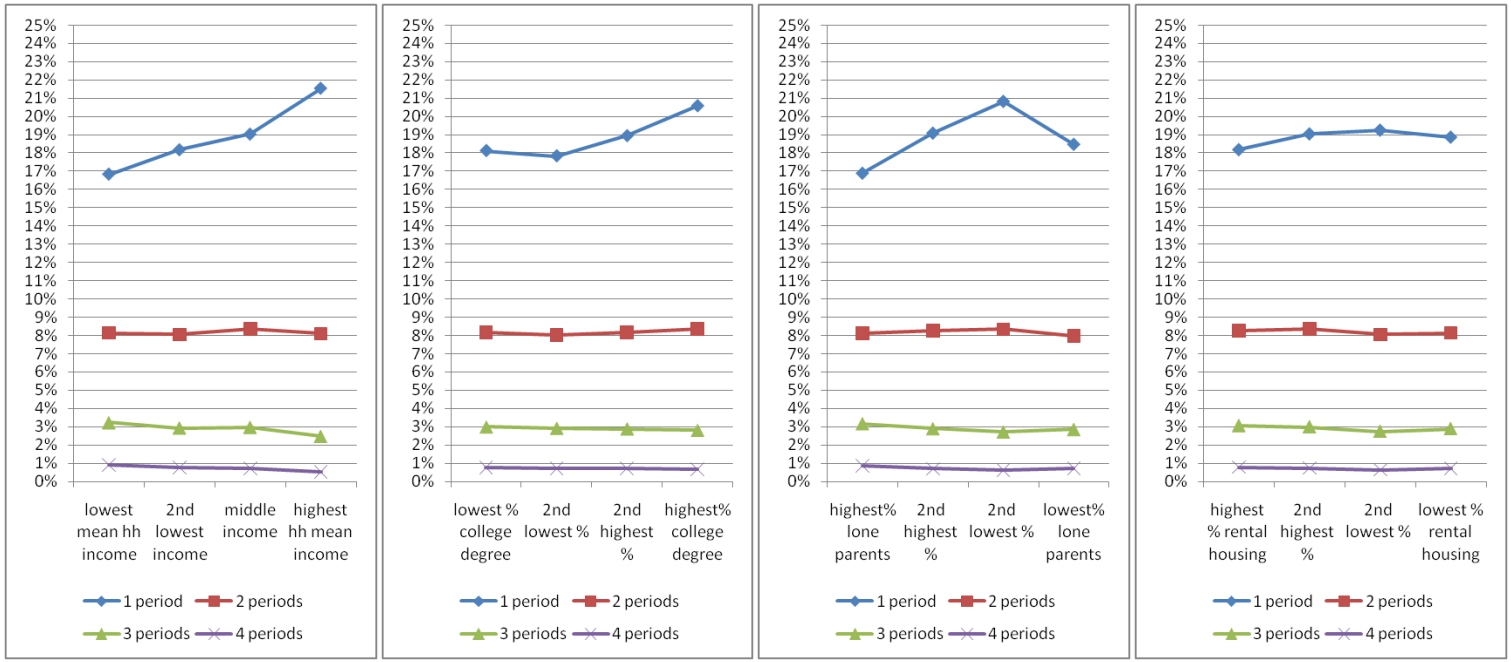
Note: horizontal axis presents highest % with lone-parents, second highest %, second lowest % and lowest % with lone-parents in neighbourhoods, respectively. Mean represents the point estimate of the transitional probability, the vertical line represents the 95% confidence interval of the point estimate.

Figure 1.A8 Predicted transitional probabilities based on random effects model by quartiles of neighborhood living arrangement



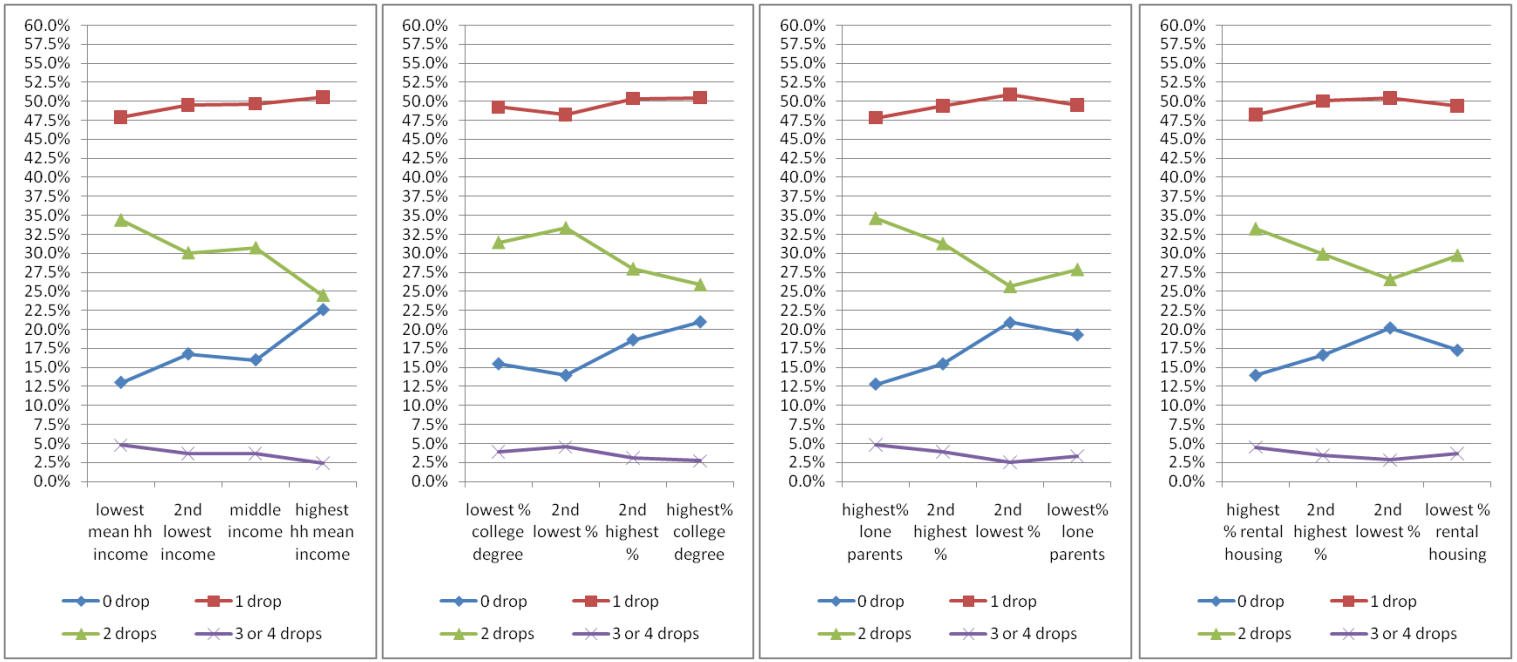
Note: horizontal axis presents highest %, second highest %, second lowest % and lowest % with rental accormodations in neighbourhoods, respectively. Mean represents the point estimate of the transitional probability, the vertical line represents the 95% confidence interval of the point estimate.

Figure 1.A9 Predicted conditional probabilities of different child health scenarios based on random effects model—the duration of a health drop at any time within six cycles by long-term neighborhood status



Note: the first panel represents the predicted conditional probabilities by quartiles of neighborhood average household income; the second panel represents the predicted conditional probabilities by quartiles of proportion of population with college degree in the neighborhood; the third panel represents the predicted conditional probabilities by quartiles of proportion of families headed by lone-parents in the neighborhood; the fourth panel represents the predicted conditional probabilities by quartiles of proportion of families living in rental accommodations in the neighborhood; respectively.

Figure 1.A10 Predicted conditional probabilities of different child health scenarios based on random effects model—the number of health drops within six cycles by long-term neighborhood status



Note: the first panel represents the predicted conditional probabilities by quartiles of neighborhood average household income; the second panel represents the predicted conditional probabilities by quartiles of proportion of population with college degree in the neighborhood; the third panel represents the predicted conditional probabilities by quartiles of proportion of families headed by lone-parents in the neighborhood; the fourth panel represents the predicted conditional probabilities by quartiles of proportion of families living in rental accommodations in the neighborhood; respectively.

Chapter 2
Family Social-economic Status, Childhood Life-events and the Dynamics of Depression from Adolescence to Early Adulthood

2.1 Introduction

The critical role of child/adolescent physical health in subsequent health and economic outcomes is well established (Case et al. 2002; Case et al. 2005; Currie 2009). Increasingly, the importance of child/adolescent mental health and behaviour problems is recognized and emphasized. One reason for this is the prevalence of child mental health problems. The MECA Study (Methodology for Epidemiology of Mental Disorders in Children and Adolescents) found that approximately one in five children and adolescents in the U.S. exhibit some impairment from a mental or behavioural disorder, 1.1 percent have significant functional impairments and 5 percent suffer extreme functional impairment (David Shaffer et al. 1996; U.S. DHHS 1999). A second reasons is that poor

mental health conditions in childhood are related to various negative consequences on future outcomes such as health status, educational attainment and labour market outcomes (Case et al. 2005). As described in Heckman's skill formation framework (Heckman 2007), health is a "capacity" that affects production of a wide range of future capacities. As Cunha and Heckman discuss, both "cognitive" ability and "non-cognitive" abilities such as perseverance, motivation, time preference and self-control have direct effects on wages, schooling, smoking, crime and many other aspects of social and economic life (Cunha and Heckman 2007; Heckman 2007).

Among all the mental health problems in the transition period of adolescence to early adulthood, depression is one of the most common (Asarnow et al. 2009). It has been estimated that 15% to 20% of youth suffer from depressive disorders by the age of 18 (Lewinsohn 2002). In the United States, 28.3% of high school students report periods of depression during the past year that interfered with usual activities and lasted at least for 2 weeks (Centers for Disease Control 2002). There is also an increasing recognition that the presence of depressive disorders often starts in the period of childhood and adolescence (Chang 2009), and depression during this transition period often persists into adulthood (Colman et al. 2007). Adolescents who experience depression often struggle with depression

throughout their lives (Lewinsohn et al. 1999), and in many cases, early onset of depression predicts more severe depression during adulthood (Weissman et al. 1999). In the period of adolescence, depression is associated with poor health and behavioral outcomes (Saluja et al. 2004), lower achievement on tests and poorer peer relationships (Roeser et al. 1998). Moreover, depression in adolescence is associated with various adverse long-term outcomes, including poor academic performance and dropping out of high school(Kessler et al. 1995; McLeod and Kaiser 2004), lower economic status, poorer labour market outcomes at later ages (Gregg and Machin 2000; Fergusson et al. 2007), drug and alcohol abuse, and suicidal behaviors (Fergusson and Woodward 2002; Fergusson et al. 2007). Detection and effective treatment of early-onset major depressive disorders can be more important than for late-onset depressive symptoms. Greden (2001) documented that early-onset depression (before the age of 21 or 22) is associated with longer first episodes, higher rates of recurrence of major depression, higher overall rates of comorbid personality disorders, and longer hospitalizations. Berndt et al. (2000) found that early-onset depression can lead to reduced educational attainment and other human capital loss, particularly for women; a randomly selected 21-year-old woman with early-onset major depressive disorder in 1995 could expect future annual earnings that were 12%-18% lower than those

of a randomly selected 21-year-old woman whose onset of major depressive disorder occurred after age 21 or not at all. Given the increasing prevalence and the negative consequences, it is of great importance to understand the evolving process of depression during the transition period of adolescence to young adulthood.

The empirical literature has documented a link between socio-economic status and depression. Among adults, depression has been shown to be associated with income in a wide variety of settings (Bruce et al. 1991; Dohrenwend et al. 1992; Murphy et al. 1991). Using an instrumental variables approach, Ettner (1996) found evidence that the association between income and depressive symptoms is causal. Moreover, unemployment has been shown to lead to depression (Rice and Miller 1995; Hamilton et al. 1997). Zimmerman and Katon (2005) found that while income loses much of its relationship to depression when other variables are controlled, employment status and financial strain are more robust predictors of depression.

Literature in psychology points out that family socio-economic status can affect the outcome of depression among adolescents: low family SES can lead to depression in adolescence transmitted by parent-child interaction patterns while high family SES can serve as a protective factor that improves resilience in youth

(see Lee and Eden 2009). There are few empirical studies that attempt to examine the relationship between family or individual SES and depression in adolescents or young adults, and the results from these empirical studies are mixed and inconclusive. Graetz (1993) showed that there is an association between unemployment and depression among Australian young men and women. Using data from the National Longitudinal Study of Adolescent Health (Add Health), Goodman et al. (2003) examined the socioeconomic status (SES) gradient on adolescents' mental health and found that the effect of income and education on depression were large. However, some empirical studies have found "no relationship" between depression among adolescents and socioeconomic status. Waschbusch et al. (2003) examined the relationship between depression and SES measured by the Hollingshead Four-Factor Index (Hollingshead 1975) in a sample of adolescents and found no association. In the examination of the trajectories of depressive symptoms among a sample of African-American youth aged 14 to 17, Repetto et al. (2004) found that depressive symptoms were not related to parental occupation. Using the National Longitudinal Study of Adolescent Health (AddHealth) data, Rushton et al. (2002) examined factors associated with persistent depressive symptoms among 13,568 adolescents who completed the initial survey in 1995 and were followed up 1 year later. They

found that socioeconomic status did not predict persistent depressive symptoms. Other studies have attempted to draw causal inference on the SES-depression gradient among adolescents. The analysis of depression from Miech et al. (1999) found no support for either causation or selection processes, suggesting that SES and depression have little influence on each other before age 21. In the Great Smoky Mountains Study, Costello et al. (2003) examined the effect of family income on children's mental health by exploiting a natural experiment involving the opening of a casino on an Indian reservation. They found that family income (especially moving out of poverty) had a positive effect on the health conditions of conduct and oppositional disorders for the children, but there was no such effect on anxiety and depression.

Our study examines the roles of family SES, early childhood life-events, unobserved heterogeneity and pure state dependence in explaining the distribution of depression among adolescents and young adults using the US data on the children of the US National Longitudinal Survey of Youth 79 (NLSY79). We employ a conditional quantile regression framework to approach this important question. Compared with conditional mean estimation models, which have been dominantly used in this literature, this approach allows us to examine the differential effects of the factors of interest at different parts of the depression

distribution, therefore providing us a more complete view of the links between these factors and youth depression. As described in the previous paragraph, some studies found statistically significant associations between family SES and youth depression while others found no such links. These discrepancies might be due to the asymmetry existing in the conditional distribution of youth depression in relation to these factors. If indeed the conditional distribution is asymmetric in nature and therefore the effects of these factors are important only at certain parts of the conditional distribution of youth depression, using a conditional mean fit may average out these effects. In addition, the conditional mean regression is often strongly affected by the behaviour of outliers in data. This non-robustness may be a potential reason for different sizes of the estimates across studies that are based on different data sets. Therefore, using a conditional quantile approach provides us with an opportunity to explore the source of discrepancies found in the existing empirical literature. To begin, we estimate a set of static conditional mean and conditional quantile models. By comparing the results from the conditional mean models with those from the conditional quantile estimation models, we attempt to investigate whether the effects of key family SES conditions and the effects of early childhood life-events vary across different quantiles.

In addition, our study explicitly models the depression dynamics from adolescence to early adulthood, an issue that has not been addressed in other studies. It is important to quantify both the mobility and persistence of this type of mental health problem over time because it helps to understand the health human capital accumulation process in the period of early life course and the protective effects of certain family SES factors in this accumulation process. Moreover, if pure state dependence of youth depression exhibits asymmetry across different quantiles of the depression distribution, this will generate policy implications that devote more resources to those with higher persistence of depression. Our study employs a newly-developed instrumental variable approach suggested by Galvao (2011) for the quantile regression dynamic panel model with fixed effects. This estimator not only allows us to control for individual-specific heterogeneity via fixed effects in the dynamic panel data framework, but also effectively reduces the dynamic bias generated by conventional dynamic fixed-effects estimation of the quantile regression models.

This paper proceeds as follows. Section 2 describes the data set we used for the study and presents some descriptive analysis of the data. Section 3 introduces the empirical methods of the study. In Section 4, the regression results are reported and analyzed. In Section 5 some conclusions are provided.

2.2 Data and Sample

2.2.1 Data Source

This study uses data on the children of the US National Longitudinal Survey of Youth 79 (NLSY79). The NLSY79 child sample is an ongoing biennial panel survey that began in 1986 and which interviewed the children born to the female respondents of the 1979 cohorts of the NLSY. Data is currently available through the twelfth wave (2008 collection). The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-competence of the children as well as the quality of their home environment (see NLSY79 Child &Young Adult Data Users Guide 2006 cycle[16]). Starting in 1994, children who reach the age of 15 by the end of the survey year are no longer assessed but instead were given the young adults survey[17] akin to that given to their mothers during late adolescence and into adulthood. This Young Adult questionnaire focuses on the transition to adulthood, with detailed questions on

[16] The most recent NLSY79 Child &Young Adult Data Users Guide is the 2006 version, but the data is available for use till the 2008 collection.

[17] As noted in the 2006 NLSY79 Child and Young Adult Data Users' Guide, there are other restrictions in the sampling procedure: "In 1994 and 1996, the Young Adult sample included all children who were age 15 and over by December 31 of that year and who met the other selection criteria. Due to budgetary constraints, the Young Adult sample in 1998 was limited to youth through age 20 as of the interview date. In 2000 only, around 40% of the black and Hispanic oversample cases between the ages of 15 and 20 were not fielded for budgetary reasons."

education, employment, training, health, family experiences, attitudes, interactions with family members, substance use, sexual activity, non-normative activities, computer use, health problems, and pro-social behaviors. As the children of the NLSY79 sample age, they continuously move into the Young Adult Interview. According to the 2006 NLSY79 Child and Young Adult data user's guide, in 1994 a total of 7,089 children who were born to the original 6,283 NLSY79 female respondents were interviewed, and among these, 6,109 were under age 15 and 980 were 15 years or older. In 2006, a total of 7,816 children, including young adults, were interviewed. Of these, 1,972 were under age 15 and 5,844 were interviewed as young adults.

These older children or "young adults" constitute the main study sample in our analyses. From the Young Adult Survey, we constructed the repeated measures of depression of these older children and other relevant variables that are potential factors determining depression in young adulthood. Drawing on the extensive information in the Child Survey, we constructed variables representing important life-course characteristics of the young adults in the period of childhood. In addition, we constructed family-level variables by using the information contained in the main NLSY79 survey, which provides more information on the mothers of the young adults. Information from the Child Survey, the Young Adult

Survey and the main NLSY79 survey can be linked by the unique identifiers of the child and the mother (seeTable A.1 in Appendix for a description of the dependent and independent variables we constructed from different sources and the corresponding questions in the survey).

2.2.2 Study Sample and Variables

2.2.2.1 Variable Definitions

The outcome variable is a scale of depression-- the Center for Epidemiological Studies Depression Scale (CES-D) developed by Radloff (Radloff 1977). The CES-D has been used in a large body of studies on depression and has been shown to have very good validity and reliability in the general population and in a wide variety of specific ethnic and socioeconomic sub-populations (Beekman et al.1997; Prescott et al.1998; Thomas et al.2001; Weissman et al.1977). The full-version of CES-D includes 20 questions related to symptoms of depression. Examples of such questions include: “In the last week I felt that I couldn’t shake off the blues, even with help from my family and friends”, and “In the last week I felt that everything I did was an effort.” Responses are coded on a scale from 0 to 3, with 0 representing “rarely/none of the time” and 3 representing “most/all of the time”. Accordingly, the composite CES-D score ranges from 0 to 60. In the

Young Adults Survey, the respondents completed a 7-item, reduced version of the CES-D questionnaire in all the cycles from 1994 to 2008. A set of seven questions was administered with skip patterns based on age and interview status. Specifically, the CES-D scale was administered to all eligible young adults in 1994 through1998, and 2004 through 2008. But in 2000, it was administered only to the eligible young adults who were not interviewed in 1998, and in 2002 it was administered only to the eligible young adults who were not interviewed in 2000. As in the full-version of CES-D questionnaire, the answers to these 7 questions were coded on a scale from 0 to 3 with 0 representing “rarely/none of the time” and 3 representing “most/all of the time”. Our study employs the 7-item composite CES-D score (ranging from 0 to 21) as our dependent variable in the analyses. From this point on, we use "the CES-D score" to represent the composite score of the 7-item questions.

In explaining the dynamics of youth depression our study focuses on family social-economic environment, prenatal or biological factors, child cognitive abilities, stressful life-events in childhood, and young adults’ own socio-economic status. A set of demographic variables for the young adults is constructed, including age, gender, race, birth order and marital status. Variables representing living environment are also included, such as, whether the youth lives in an urban

or rural area, and whether the youth lives in a Standard Metropolitan Statistical Area (SMSA). We include a set of biological factors including age of mother at the birth of the child, mother's drinking, smoking and substance use one year prior to the birth of the child.

In the psychology literature, experience of traumatic life-events has been identified as one of the most important risk factors associated with elevated risk of depression (Lee and Eden 2009). In the Child Survey, a question was asked about whether the child had a psychological consultation in the previous 12 months; if the answer is "Yes", the respondent was asked the reason for the consultation. We use two variables to capture traumatic life experiences in the period of childhood: whether a child consulted a psychiatrist in the previous 12 months due to emotional trauma, molestation or abuse, and whether the child consulted a psychiatrist because of loss of parents/siblings or divorce of parents. Although we observe repeated measures of these two variables over multiple cycles, we constructed variables measuring the number of times in the past that a child consulted a psychiatrist because of these two problems.

To capture socio-economic factors we consider both the parental socio-economic variables and the young adults' own SES because they may have different effects on the dynamics of depression during adolescence to young

adulthood. We measured maternal education as the highest grade completed by the mother. We include measures of maternal employment status: the variables are defined as the number of weeks unemployed in the past calendar year and the number of weeks unemployed since the last interviewWe constructed a parental income measure as the total net family income in the family of the mother, which is included in the Main NLSY79 Survey. It worth noting that this variable will be missing if the young adult was living in the father's or another relative's household at the time of the Young Adult Interview.[18] The only employment measures of the youth administered consistently in the Young Adults Interview relate to a young adult's "significant job" defined as the last job lasting two weeks or more in the last year[19].

2.2.2.2 Sample Definition

The total sample of individuals who ever completed a Young Adult Survey during the survey years of 1994-2008 is 7,100. We used several criteria to select our sample. First, we only kept the individuals in the Youth Survey who had at least

[18] Only since 2000 is there a question in the Young Adult Interview asking the total family income of the respondents, which refers to the income from all sources by all the family members. Because we don't have such a measure for the previous cycles, we do not use this family income measure for our study.

[19] We also considered other family and youth SES factors, including highest grade completed by the father, paternal unemployment status, young adults' own education variables such as year of school currently enrolled in, highest grade of regular school completed, and whether the respondent ever repeated or skipped grade, and young adults' own income. But due to a large proportion of missing values, these variables are dropped from our estimation analyses.

one wave of observation of the CES-D score during the survey years of 1994-2008. Imposing this criterion excludes 20 individuals, reducing the available sample to 7,080 individuals. Second, we dropped the observations for which an individual was aged 26 or above in any wave of the Young Adults Interview, because the focus of our study is the dynamics of depression during adolescence to young adulthood. This leads to a further reduction of the sample to 7,035 individuals. Third, we dropped individuals with fewer than three consecutive waves of observation of the CES-D score, because we need to include the first lag of the CES-D score to estimate a dynamic model and the second lag of the dependent variable as the instrumental variable for the IV approach we employ for the conditional quantile estimation (details see section 3.2.6). After applying this criterion, we have 3,543 individuals with 11,558 observations in total as our study sample.

2.2.3 Descriptive Statistics of the Study Sample

In Table 1, we list the summary statistics of the variables we use for the estimation models across all individuals in our study sample and over all waves[20]. As noted

[20] In order to estimate the dynamic models with relatively more time periods, we decided to drop some of the variables with a large proportion of missing values from the estimations. Therefore, only a subset of the variables described in the previous section are included in for the descriptive

above, we only include observations for young adults under age 26. In this sample, the mean age of all of observations over time is around 19 years old and about half of the individuals are male. The CES-D depression score has a mean of 4.58 and a standard deviation of 3.68. There are a relatively large number of zero scores in the sample—about 11.2% of the observations have CES-D scores of zero, meaning no symptoms of depression. Figure 1 presents a histogram of the CES-D score for our full analytic sample. The distribution of the CES-D score has a long right tail, with more than 95% of the values under 12. (A more detailed breakdown of the distribution of the CES-D score is presented in Table A.2 in the Appendix.)

In order to describe the transitions among different levels of depression, Table 2 presents the transition matrix for the CES-D score classified into five categories: score 0, score 1-3, score 4-6, score 7-11 and score 12 and above. The categories are chosen according to some typical values in the CES-D distribution (i.e., they are not based on clinical classifications). The rows of the transition matrix indicate the depression level in the previous period, while the columns indicate the depression level in the current period. The transition matrix shows that the majority of the transitions among different levels of depression appears on the

statistics in Table 1 and for all the following regression analyses.

diagonal or one cell off the diagonal. This indicates that persistence exists in the dynamics of depression for the young adults, and the most persistence is observed for those with CES-D scores of 1-3 or 4-6, suggesting a potential benefit from using quantile regression models for depression dynamics.

2.3 Empirical Methods

At first, we estimate a series of static conditional mean and conditional quantile models to examine the roles of family SES, childhood stressful life-events, prenatal and biological factors in explaining the distribution of youth depression. We then use a series of dynamic conditional mean and conditional quantile models to examine the dynamics of depression during adolescence to young adulthood. The way we capture the dynamics is to control for the first lag of depression score, in addition to all the covariates in the static models. For simplicity we only describe the methodological issues and the empirical specifications for the dynamic models in the following discussion. The only difference in the specifications between static and dynamics models is that in static models the first lag of depression score is not controlled for as one of the regressors.

2.3.1 Quantile Regression Dynamic Panel Instrumental Variable Model with Fixed Effects

Our study employs an instrumental variable approach suggested in Galvao (2011) for a quantile regression dynamic panel model with fixed effects. The method is adapted based on the instrumental variables quantile regression method of Chernozhukov and Hansen (2005, 2006, 2008) along with lagged regressors as instruments. This estimator provides us several advantages for the analysis of depression dynamics. First, it is important to separate individual-specific heterogeneity from state dependence in the context of studying the persistence of health outcomes (see Contoyannis 2004a, 2004b); this estimator allows the control of individual-specific effects via fixed effects in the dynamic panel data framework. Second, exploring heterogeneous covariate effects within the quantile regression framework offers a more flexible approach than the classical Gaussian fixed- and random-effects estimators (Galvao 2011). In our context, we use this method to explore potential heterogeneity in the effects of factors of interest on the dynamics of depression across the quantiles. Third, the quantile regression model has a significant advantage over models based on the conditional mean, since it will be less sensitive to observations in the tail of the underlying random variables, and consequently will be less sensitive to outliers. This approach can

provide robust estimates that do not rely on specific assumptions of the outcome distributions. Fourth, the IV-estimator reduces the estimation bias generated by the conventional fixed-effects estimation of the quantile regression. Galvao (2011) shows that under some mild regularity conditions (notably with $T \rightarrow \infty$ as $N \rightarrow \infty$ and $N^{\alpha}/T \rightarrow 0$, for some a>0), the estimator is consistent and asymptotically normal. More importantly, Monte Carlo experiments show that even in short panels this instrumental-variable estimator can substantially reduce the bias. As in most of the micro-level panel data cases, our data is a short panel of large cross-sections (large N and modest T). Therefore, this property is important in generating estimates of our dynamic model.

In a dynamic model for panel data with individual fixed effects, the τth conditional quantile function of the outcome variable of the tth observation on the ith individual y_{it} can be represented as

$$Q_{yit}(\tau|z_i, y_{it-1}, x_{it}) = z_i\eta + \alpha(\tau)y_{it-1} + x'_{it}\beta(\tau), \qquad (1)$$

where y_{it} is the outcome of interest, y_{it-1} is the lag of the variable of interest, x_{it} are a set of exogenous variables, z_i is an individual identifier, and η represents the Nx1 vector of individual specific effects. Since it is difficult to estimate a τ-dependent distributional individual effect in a short panel of large cross-sections (large N and modest T), Galvao (2011) restricts the estimates of the individual

specific effects to be independent of τ across the quantiles and estimates the model for several quantiles simultaneously. In other words, only the effects of the covariates (y_{it-1}, x_{it}) are allowed to depend on the quantile τ of interest in the above model. Koenker (2004) introduced a general approach to the estimation of quantile regression fixed-effects models for panel data. Galvao (2011) applied this general approach to a dynamic panel model estimation, and proposed an estimator which solves

$$\left(\hat{\eta}, \hat{\alpha}, \hat{\beta}\right) = \min_{\eta,\alpha,\beta} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \upsilon_k \rho_\tau \left(y_{it} - z_i\eta - \alpha(\tau_k) y_{it-1} - x_{it}^{'}\beta(\tau_k)\right), \quad (2)$$

where $\rho_\tau(u) \coloneqq u(\tau - I(u < 0))$ as in Koenker and Bassett (1978), and υ_k are the weights that control the relative influence of the K quantiles $\{\tau_1, \ldots, \tau_K\}$ on the estimation of the η_i parameters.

The quantile regression fixed-effects estimator based on Equation (2) suffers from bias in the presence of lagged dependent variables as regressors. Using an analogous rationale for the construction of instruments as in Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991), Galvao (2011) suggests that valid instruments that can be used to produce a consistent estimator for dynamic panel data models are available from inside the model. Specifically, because the lagged (or lagged differences of) regressors are correlated with the included regressors (lag of the dependent variable in our case) but are uncorrelated with the error term,

they can be used as valid instruments. Following Chernozhukov and Hansen (2006, 2008), Galvao (2011) then proposed an IV estimator for the state dependence parameter.

The implementation of this quantile regression instrumental variables procedure requires minimizing the objective function

$$Q_{NT}(\tau, \eta_i, \alpha, \beta, \gamma) := \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \upsilon_k \rho_\tau \left(y_{it} - z_i \eta - \alpha(\tau_k) y_{it-1} - x_{it}' \beta(\tau_k) - \omega_{it}' \gamma(\tau_k) \right), \quad (3)$$

where y_{it-1} is a dim(α)-vector of endogenous variables, z_i identifies the individual fixed effects, x_{it} is a vector of exogenous explanatory variables, ω_{it} is a *dim(*γ*)*-vector of instrumental variables such that *dim(*γ*) >=dim(*α*)*. Specifically, the instrumental variables may include values of *y* lagged two periods or more and/or lags of the exogenous variable *x* which affect the determination of lagged y but are independent of *u*. The author then suggests using a numerical optimization function in R to implement this estimator, which minimizes the coefficient of the instrumental variable in the same problem as described above (Equation (2)). The intuition is that if the instrument is valid, it is independent of the error term and should have a zero coefficient.

A further complication in our empirical model comes from the fact that the outcome variable is an ordered discrete response—the CES-D score. In this case, the above conditional quantile regression model (used for continuous outcome

variables) may be problematic because the cumulative distribution function of the CES-D score is discontinuous with discrete jumps between flat sections, so the quantiles of this discrete variables are not unique[21]. As noted by Machado and Santos Silva (2005), the main problem with estimating conditional quantiles for discrete responses(e.g. count data) stems from the conjunction of a non-differentiable sample objective function with a discrete dependent variable. To extend the conditional quantile regression to count data, Machado and Santos Silva (2005) proposed an approach which adds artificial smoothness to the data using a form of "jittering process". Specifically, the artificial smoothing is achieved by adding uniformly distributed noise to the count variable. This way they construct a continuous variable with conditional quantiles that have a one-to-one relationship with the conditional quantiles of the original counts. Then this artificially constructed continuous variable is used as the base for inference. Machado and Santos Silva (2005) show that this approach of smoothing allows inference to be performed using standard quantile regression techniques. However, a problem of this approach is that it introduces extra noise to the quantile regression estimators. To reduce the effect of this unnecessary noise, the parameters of the model are estimated multiple times using independent draws

[21] By convention, the lower boundary of the interval defines the quantile in such a case in most of the current literature.

from the uniform distribution, and the multiple estimated coefficients and confidence intervals are averaged over the jittered replications. We followed their implementation of the jittering process, and then estimated the above instrumental variable with fixed-effects model with the jittered data.

2.3.2 Empirical Specifications and Estimation Methods

We examine the level of state dependence of the CES-D score and the inter-temporal roles of family SES, childhood stressful life-events, prenatal and biological factors in explaining the distribution of youth depression using both the conditional mean and conditional quantile estimation models. The empirical specifications of these dynamic panel data models are described in the sub-sections below[22].

2.3.2.1 Pooled dynamic conditional mean estimation models

For a conditional mean estimation without considering individual heterogeneity, we consider the following specification:

$$E(y_{it}|y_{it-1}, x_{it}) = \alpha y_{it-1} + x_{it}'\beta, \quad (4)$$

where y_{it} is the CES-D score, y_{it-1} is the first lag of the CES-D score, x_{it} is a vector

[22] The empirical specifications for the static models are the same with those for the dynamic models except that the term of y_{it-1} is not included on the right-hand side of the equations.

of explanatory variables, including youth demographic characteristics, childhood life-events, prenatal or biological factors, and family SES characteristics. First, we estimate a pooled linear model (treating the CES-D score as a continuous variable) as a baseline regression model. To account for the ordered discrete nature of the CES-D score, we then estimate the conditional mean of the dependent variable with a pooled Poisson model. It is worth noting that for the Poisson estimation the conditional mean is not as specified in equation (4), but follows the standard parameterization of $E(y|\boldsymbol{x})=\exp(\boldsymbol{x'\beta})$.

2.3.2.2 Dynamic conditional mean models with individual-specific effects

To separate state dependence from unobserved individual heterogeneity, we consider the conditional mean estimation including individual-specific effects with the following specification:

$$E(y_{it}|y_{it-1}, x_{it}, z_i) = \alpha y_{it-1} + x_{it}'\beta + z_i\eta, \qquad (5)$$

where η denote the individual fixed effects. Again we estimate both linear and Poisson specifications of the model: we first estimate the linear random-effects and fixed-effects models, and then estimate the Poisson model with random-effects and fixed-effects specifications[23].

[23] Again for the Poisson estimation the conditional mean is not as specified in equation (5), but follows the standard parameterization of $E(y|\boldsymbol{x})=\exp(\boldsymbol{x'\beta})$.

2.3.2.3 Dynamic panel conditional quantile estimation model without fixed effects

As a baseline for conditional quantile estimation, we consider a dynamic model for the τth conditional quantile function of the outcome variable with the following specification:

$$Q_{y_{it}}(\tau|y_{it-1}, x_{it}) = \alpha(\tau)y_{it-1} + x_{it}'\beta(\tau), \quad (6)$$

where y_{it} is the CES-D score, $y_{it-1}$is the first lag of the CES-D score, x_{it} is vector of explanatory variables, including youth demographic characteristics, childhood life-events, prenatal or biological factors, and family SES characteristics. The parameter α captures the state dependence level of the CES-D score. It should be noted that all the parameters α and β in this model are allowed to depend on the quantile τ of interest.

2.3.2.4 Dynamic panel quantile estimation with jittering but without fixed effects

As noted above, the regular conditional quantile estimation for continuous data may be problematic in our context because our dependent variable is ordered and discrete. Following Machado and Santos Silva (2005), we first add randomness to our dependent variable by "jittering" the CES-D score, and then apply the jittered sample to the dynamic conditional quantile estimation. Specifically, we

replace the discrete CES-D score y_{it} with a continuous variable $J_{it}= h(y_{it})$, where $h(.)$ is a smooth continuous transformation. The transformation used is

$$J_{it}= y_{it} +u, \qquad (7)$$

where u~ U(0, 1) is a random draw from the uniform distribution on (0, 1)[24]. To allow for the exponentiation for our outcome variable (as for count data), the conditional quantile of $Q_J(\tau|X)$ is specified to be

$$Q_J(\tau|X) = \tau + exp\left(X'\beta(\tau)\right), \quad (8)$$

where X represents the design matrix in the specification of y_{it} considered in (6). The additional term τ appears in the equation because $Q_J(\tau|X)$ is bounded from below by τ. To estimate the parameters of a quantile model in the usual linear form, a log transformation is applied so that $ln(J- \tau)$ is modeled, with the adjustment that if $J- \tau <0$, then $ln(\varepsilon)$ is used, where ε is a small positive number[25]. To reduce the effect of noise due to jittering, the parameters of the model need to be estimated multiple times based on multiple jittered replications. In our study, we chose 500 jittering replications to derive the estimates for the quantile regression models[26].

---

[24] We use a random draw from the uniform distribution, following Machado and Santos Silva (2005).

[25] The log transformation with the adjustment is justified by the property that quantiles are equivariant to monotonic transformation and the property that quantiles above the censoring point are not affected by censoring from below (details see Cameron and Trivedi (2009)).

[26] We experimented with the number of jittering replications, including 50, 100, 500 and 1,000 jittered samples. We chose 500 jittered samples because increasing from 500 to 1000, the

2.3.2.5 Dynamic panel quantile estimation model with fixed effects

To account for potential unobserved individual heterogeneity, we consider a dynamic panel quantile regression with individual fixed-effects:

$$Q_{y_{it}}(\tau|z_i, y_{it-1}, x_{it}) = z_i\eta + \alpha(\tau)y_{it-1} + x_{it}'\beta(\tau), \qquad (9)$$

where z_i identifies the individual fixed effects. The estimation of the above model is implemented by a regularization method developed by Koenker (2004). Because the asymptotic inference is problematic with this estimator, we use bootstrap techniques to derive numerically the standard errors and confidence intervals for this model.

2.3.2.6 Dynamic panel instrumental variable quantile regression with fixed effects

As noted above, the instrumental variable approach suggested by Galvao (2011) can reduce estimation bias in the above dynamic panel quantile regression with fixed effects. Specifically, the estimates of the parameters are derived to minimize the objective function described by Equation (3). In our study, we use the values of CES-D score lagged two periods as our instrument.

It should be noted that the asymptotic variance-covariance matrix of the

estimation results do not change much (the difference is only to the 3rd decimal place) but the calculation time almost doubles.

above quantile regression estimator depends on the density of the error term and is not easy to compute. One option is to estimate the variance-covariance matrix directly using nonparametric techniques. However, because this estimation of the variance-covariance matrix is computationally intensive and not easily attainable, we follow the bootstrap inference approach that is used extensively for quantile regressions to derive numerically the standard errors and confidence intervals for this estimator. Specifically, we construct the bootstrap samples by resampling only from the cross-sectional units (individual persons in our case) with replacement, because Monte Carlo simulations on bootstrap inference for quantile regression for panel data suggest that cross-sectional bootstrapping has the best performance among the three different procedures of bootstrap in this context (Galvao and Montes-Rojas 2009; Kato, Galvao and Montes-Rojas 2010). We used 499 bootstrap replications with a pair-wise resampling technique to construct the empirical distribution of the estimator and construct the bootstrap standard errors. We also used a percentile bootstrap procedure to construct 95% confidence intervals for the parameters of interest.

2.3.2.7 Dynamic panel instrumental variable quantile regression with jittering jittering and fixed effects

To account for the problems arising with quantile regression with count data as

the dependent variable, we apply the above instrumental variable approach to a jittered sample. We use the same process to construct the jittered samples as for the dynamic panel quantile estimation described in Section 3.2.4. We then estimate the above IV estimator with the artificially smoothed CES-D score as the dependent variable[27]. Again we chose 500 jittering replications to derive the estimates for this IV approach quantile regression with fixed-effects model.

Since the jittering process involves a non-linear transformation from the original CES-D score to a smoothed variable, the marginal effect (ME) estimates are different from the coefficient estimates. We use the marginal effects at the mean (MEM) convention to calculate the MEs. According to Equation (8), the MEs for any continuous regressor x_j are estimated by $exp(\overline{\boldsymbol{X}}'\widehat{\boldsymbol{\beta}})\hat{\beta}_j$, with all the regressors evaluated at their mean values. For any dummy variable x_j, we calculate the MEs with respect to a change in this dummy variable from 0 to 1, using the difference of the corresponding predicted values: $exp(\overline{\boldsymbol{X_1}}'\widehat{\boldsymbol{\beta}}) - exp(\overline{\boldsymbol{X_0}}'\widehat{\boldsymbol{\beta}})$, where $\overline{\boldsymbol{X_1}}$ represents the design matrix evaluated at 1 for this dummy variable x_j and at the means for all the other regressors, while $\overline{\boldsymbol{X_0}}$ represents the design matrix evaluated at 0 for this dummy variable x_j and at the

[27] When we estimate the quantile models with jittering (as in Sections 3.2.4 and 3.2.7), we are estimating marginal effects for a different specification (conditional quantile is an exponential function of X) than when we are estimating the conditional quantiles assuming continuity (as in Sections 3.2.3, 3.2.5 and 3.2.6), where conditional quantile is specified as a linear function of X.

means for all the other regressors.

2.4 Estimation Results

2.4.1 Results for Conditional Mean Estimation with Static Models

Table 3 presents results for the conditional mean estimation for the CES-D score based on the static linear panel data models. Columns (1) and (2) present marginal effects and standard errors for the pooled linear model; columns (3) and (4) present the results for the random-effects model; and columns (5) and (6) present the results for the fixed-effects model. Several patterns can be observed from the results. First, as indicated in the current literature on youth depression, demographic characteristics are important in explaining the variability of depression. Females and blacks are more susceptible to depression. Birth order seems important as well: those born later in the family are more likely to experience depression. Second, there is a statistically significant and large positive correlation between psychological health care utilization and the presence of depression. Surprisingly, the pooled model suggests that the incidence of family problem and emotional trauma are not statistically significant in explaining the variability of youth depression. This contradicts with the evidence in this literature indicating an adverse effect of early-life stressful life events on youth depression.

But the results based on the random effects model show that, after taking into account the individual heterogeneity, both types of stressful life events (family member loss or emotional trauma) are significantly associated with higher CES-D scores in young adulthood,. Fourth, prenatal factors including the age of mother at birth of the child, maternal drinking and smoking behaviours are statistically significant in the model: the children who were born to younger mothers, and those born to mothers with drinking or smoking behaviours during pregnancy are likely to have higher depression scores. Lastly, among the set of family SES factors, maternal education and family poverty status are important in explaining the variability of youth depression: lower maternal education and living in poverty are associated with higher CES-D scores. On the other hand, maternal unemployment status and total family income are not associated with youth depression.

Table 4 summarizes the results for the pooled model, the random-effects model and the fixed-effects model using Poisson specifications. The reported standard errors for the random- and fixed- effects models are based on bootstrapping for 499 replications. The patterns found in the linear model regressions are preserved in the Poisson estimations. The sizes of the estimated marginal effects are also similar to those based on pooled model. This indicates

that a linear specification is appropriate to model the conditional mean of the CES-D score with our data, although the CES-D score exhibits nonlinearity and discreteness as described by the descriptive statistics and the histogram of the CES-D scores.

2.4.2 Results for Conditional Quantile Estimation with Static Models

Table 5 presents the static conditional quantile estimation results for the pooled model. Columns (1) and (2) list the marginal effects and the standard errors for the estimation of the 0.25 conditional quantile of the CSE-D score; Columns (3) and (4) list the results for the estimation of the 0.5 conditional quantile of the CSE-D score; Columns (5) and (6) present the results for the estimation of the 0.75 conditional quantile of the CSE-D score. The signs of the estimated marginal effects are consistent with those based on the pooled conditional mean estimation model, except for the two geographical variables flagging youth living in urban area and youth living in SMSAs. But these two variables are statistically insignificant in both models, so reversed signs is not indicative of differential directions of the effects. The magnitudes of the marginal effects in general vary across different quantiles and for some of the variables a clear gradient is observed. First, gender difference of youth depression is larger at the higher end

of the depression distribution. Second, the positive association between emotional problem consultation or drug use for behaviour problem and youth depression is stronger at the higher end of the distribution. Third, prenatal and biological factors have different roles at different points of the depression distribution. The positive link between higher depression score and maternal drinking and smoking behaviours is stronger at the higher ends of the CES-D score distribution; as a matter of fact, maternal drinking behaviour is not statistically significant at the lowest quantile of the CES-D distribution. The link between higher depression score and younger maternal age at the birth of child is stronger at the higher ends of the distribution. Interestingly, stressful life-events during childhood are statistically significant only at some of the quantiles and different types of events play different roles across quantiles. The incidence of family problems during childhood, i.e. parents divorce or family member loss, plays a more important role at the higher ends of the CES-D distribution, while the incidence of emotional trauma during childhood, i.e. abuse or molestation, seem only to contribute to the variability of depression for the individuals who have relatively low-severity depression. This reveals the asymmetry of the link between early-life events and youth depression. Studies based on conditional mean estimation may neglect this aspect of the link thus provide very different implications. Finally, the roles of

family SES characteristics differ across different quantiles of the CES-D distribution. The link between higher depression and lower family SES, i.e. lower maternal education and family being in poverty, is stronger at the higher ends of the distribution, highlighting the critical role of these three factors for individuals who have more severe depression levels.

Table 6 summarizes the conditional quantile estimation results for the pooled model without individual-specific effects based on the jittered sample. Again the marginal effects and the standard errors are presented separately for the 25th quantile, the 50th quantile and the 75th quantile of the CSE-D score. The estimates presented in the table are based on 500 jittering replications. The general patterns observed in the previous estimation are mostly preserved in this model. However, additional insight can be obtained from the results based on this model. In addition to maternal education and family poverty status, estimates also indicate the non-trivial role of maternal unemployment. The variable "maternal number of weeks unemployed in the last year" becomes statistically significant at the highest quantile of the CES-D distribution. Moreover, the results show a clearer family SES-youth depression gradient: the associations between lower youth depression and higher maternal education and maternal unemployment is stronger at the higher ends of the CES-D distribution. This reveals an asymmetric

nature of the family SES-youth depression link, and again stresses the more important roles of family SES factors for the individuals who experience more severe levels of depression. The additional insight we obtain from the results based on jittered sample illustrate the advantage of using the smoothing technique proposed by Machado and Santos Silva (2005) to model the conditional quantiles of count data. Since standard inference methods are asymptotically valid for this model, estimates based on jittered sample provide us with a more accurate view of the family SES and youth depression associations.

2.4.3 Results for Conditional Mean Estimation with Dynamic Models

Table 7 summarizes the conditional mean estimation for the CES-D score based on the linear dynamic panel data model. Columns (1) and (2) present marginal effects and standard errors for the pooled linear model; columns (3) and (4) present the results for the random-effects model; and columns (5) and (6) present the results for the fixed-effects model. The estimated marginal effect of the first lag of CES-D score captures the pure state dependence of youth depression conditional on all the other covariates we have discussed in the static models. Both the pooled and the random-effects models indicate a strong positive state dependence of depression during adolescence to early adulthood: the CES-D score

in the current period is positively correlated with that in the previous period. However, the state dependence estimate based on the fixed-effects model is statistically significant and negative, suggesting that the positive correlation between the previous depression score and the current depression score disappears after controlling for the unobserved individual heterogeneity. The intuition may be that conditional on all the other variables, the variation in depression scores is mainly mean-deviation after taking into account the individual fixed effects according to the conditional mean estimation model. It is worth noting that the magnitude of the estimated persistence level based on the fixed effects model is surprisingly large, but this estimate is subject to dynamic bias and thus needs to be interpreted with caution. The intra-class correlation coefficient (ICC) estimate suggests that about 21.2% of the error variance is attributable to unobserved heterogeneity in the random-effects model, while about 70.5% of the error variance is due to the unobserved individual heterogeneity in the fixed-effects model.

Several patterns are observed about the inter-temporal effects of other covariates on youth depression. First, the signs of the marginal effects are the same with those from the static models. This indicates that the associations between the factors of interest and youth depression in the long run are preserved

in the transitional process. The dynamic model results are in line with the static models in the following ways. The pooled model and random-effects model results indicate that youth depression varies substantially across different demographic characteristics: females and blacks are more likely to experience depression; the youth who were born later in the family are more likely to be depressed. There is a positive correlation between psychological health care utilization and the presence of depression. Young adults who utilize consultations for emotional or behaviour problems, and who take prescription drugs to control their activity level, tend to have higher depression scores. Family SES characteristics including maternal education and family poverty status, are important in explaining youth depression: higher maternal education is associated with lower depression scores; adolescents who are brought up in a low-income family or a deprived family are more likely to be depressed. Second, the magnitudes of the marginal effects in the dynamic models are in general smaller than those in the static models based on pooled and random-effects specifications. In fact, the prenatal factors become statistically insignificant in the dynamic models. This makes sense because the results from the dynamic model only capture the inter-temporal effects of these factors conditional on the previous depression. The only exception is with the incidence of family events: the

estimates from the dynamic models become statistically significant and are slightly larger than those from the static models. According to both the pooled and the random-effect models, the incidence of emotional trauma and the incidence of family problems or loss of family members during the period of childhood are important in explaining the dynamics of depression during adolescence to early adulthood. Emotional trauma, molestation and abuse during childhood have a more serious adverse effect than family problems like divorce of parents or loss of family members. Lastly, the fixed effects model results highlight the important roles of maternal unemployment and youth living in an SMSA in explaining the dynamics of youth depression. While maternal unemployment status is found statistically insignificant in the static models, longer duration of maternal unemployment is associated with higher depression scores based on the dynamic fixed effects model. Also, youth living in an SMSA is now found to be associated with higher youth depression.

To account for the ordered discrete nature of the CES-D score, we estimate the conditional mean of the dependent variable using a Poisson specification. Table 8 lists the results from the pooled model, the random-effects model and the fixed-effects model, respectively. The reported standard errors for the random- and fixed- effects models are based on bootstrapping for 499 replications. The

patterns found in the above linear dynamic models are preserved in the Poisson estimations with some exceptions. First, the estimated state dependence level is substantially different in the linear and the Poisson specifications. With a Poisson specification, the state dependence became negative but not statistically significant in the random-effects model, and the estimate in the fixed effects model is still negative but becomes closer to zero. Second, the positive association between the psychological health care utilization and the presence of depression became statistically insignificant in the fixed-effects model. Third, the link between biological factors or prenatal factors and depression became statistically significant: older maternal age at the birth of child is associated with slightly lower CES-D scores; the incidence of maternal drinking and smoking behaviour during pregnancy is associated with higher CES-D scores. Since the CES-D score exhibits common features of count data, the estimates based on Poisson specification are more reliable than those from the linear specifications.

2.4.4 Results for Conditional Quantile Estimation with Dynamic Models

Table 9 presents the dynamic conditional quantile estimation results for the pooled model without individual-specific effects. Columns (1) and (2) list the marginal effects and the standard errors for the estimation of the 0.25 conditional quantile

of the CSE-D score; Columns (3) and (4) list the results for the estimation of the 0.5 conditional quantile of the CSE-D score; Columns (5) and (6) present the results for the estimation of the 0.75 conditional quantile of the CSE-D score. The dynamics of depression differs across different quantiles of the CES-D score in a number of ways. First, the pure state dependence of youth depression differs at different parts of the CES-D distribution. Since a higher CES-D score indicates a more severe depression symptom, the difference among the marginal effects of the CES-D lag across different quantiles suggests that the positive state dependence is stronger for the individuals with more severe symptoms of depression. In other words, it is more difficult for the young adults with worse depression to recover. Second, the inter-temporal effects of some covariates vary across different quantiles of youth depression. Consistent with the results in the static models, gender difference of youth depression is bigger at the higher end of the depression distribution, while racial difference in depression is smaller at the higher end of the distribution. According to the dynamic model, stressful life-events during childhood play a more important role at the lower end of the CES-D score. The adverse effects of family problems or family member loss, and of emotional trauma during childhood seem only to contribute to the dynamics of depression for the individuals who have relatively low severity level of depression. This

pattern is different from the one based on the static models, which suggests that the incidence of family problem during childhood plays a more important role at the higher ends of the CES-D distribution. Maternal smoking behaviour only adversely affects the individuals who have low severity of depression. This pattern is again different from the one observed in the static models. Finally, the roles of family SES characteristics differ across different quantiles of the CES-D score. The protective effect of higher maternal education is more significant and important for individuals who have more severe problems of depression. Compared with the results from the static models, the dynamic model results highlight the important role of total family income rather than the family poverty status in the dynamics of youth depression. According to the dynamic model, the inter-temporal effect of family poverty status is statistically insignificant while the total family income is statistically significant at the highest quantile of the CES-D distribution.

In order to illustrate the differences in the marginal effects across different quantiles, we present the above estimates via a graphical display of coefficients and the respective confidence intervals in Figure 2. In the figure, each separate graph presents the estimated coefficients and the 95% confidence intervals for each regressor. In each graph, the horizontal dashed lines are the pooled OLS

estimates of the point estimate and the 95% confidence interval (corresponding to estimates presented in columns 1 and 2 in Table 7). The green solid lines and the shaded areas represent the quantile regression estimates of the coefficient and the 95% confidence intervals. The first graph clearly shows that the state dependence level of depression varies dramatically at different quantiles of the CES-D score distribution: the persistence level of depression increases from the lower quantiles to higher quantiles. This pattern is not captured by the pooled mean estimation model. The confidence intervals of the quantile regressions widen at the upper quantiles, indicating that the estimates at the upper quantiles are less precise than those at the lower quantiles.

Table 10 presents the conditional quantile estimation results for the pooled model without individual-specific effects based on the jittered sample. The estimates presented in the table are based on 500 jittering replications. The general patterns observed in the previous estimation, which treats the CES-D score as continuous data, are preserved in the estimation based on artificially smoothed CES-D score. It is still observed that the estimated persistence level is stronger at the higher ends of the CES-D distribution. However, the magnitudes of the estimates based on jittered sample are slightly smaller.

Table 11 summarizes the results for the conditional quantile estimation with

individual fixed effects. Since now the individual fixed effects are added, only time-varying regressors are included in the estimation. The reported standard errors are based on 499 bootstrap replications. As in the conditional mean estimation with individual fixed effects models (as shown in Table 7 and Table 8), the estimated marginal effects on the lag of depression score are statistically significant and negative across all three quantiles. Given that the state dependence estimates from the conditional quantile models without individual fixed effects are statistically significant and positive, this again suggests that conditional on all the other variables, the variation in depression scores is mainly mean-deviation after taking into account the individual fixed effects. The absolute value of the estimated state dependence decreases at higher quantiles of the CES-D score distribution. This might indicate that there is a lower level of mean reversion for the group of individuals at higher ends of the CES-D score distribution. It is worth noting that these estimates of state dependence suffer from the dynamic bias in the dynamic quantile regression model described in the methods section. So it is still difficult to infer too much about the true state dependence level of youth depression from these estimates. Compare to the pooled dynamic quantile regression model results, some different patterns are observed. First, the consultations for emotional problems and the use of drugs for behaviour problems

are still positively associated with higher CES-D scores, but less statistically significant after controlling for individual fixed effects. Having a job is positively associated with higher depression scores. Interestingly, after controlling for the individual fixed effects, maternal unemployment duration becomes statistically significant at the higher ends of the CES-D distribution; while maternal education becomes statistically insignificant. The positive link between poverty status and higher youth depression scores is observed across all the quantiles but remains statistically insignificant as in the previous conditional quantile estimation models without individual fixed effects.

Table 12 presents the results for the instrumental variable conditional quantile estimation with individual fixed effects. The estimation is based on the original CES-D score without the jittering process. The reported standard errors are based on 499 bootstrap replications. After instrumenting the first lag of CES-D score, the estimates for the persistence level change dramatically across all the quantiles. The estimated state dependence parameter becomes statistically insignificant across all quantiles and the magnitudes of the estimates are substantially reduced, becoming much closer to zero. Since the IV estimator can reduce the dynamic bias of the state dependence parameter, these estimates should be more reliable than those from the model without instrumenting. The fact that the estimates are

statistically insignificant and close to zero indicates that the persistence level of youth depression is very small at all parts of the CES-D distribution. The association between the previous CES-D score and the current CES-D score is merely mean deviation around the within-individual means after taking into account the unobserved individual heterogeneity. It is worth noting that after instrumenting the first lag of CES-D score with the second lag of the CES-D score, we lost a wave of observations. This, in conjunction with the use of instrumental variables, increases the standard errors dramatically: the bootstrapped standard errors are at least twice those based on individual fixed effects without instrumenting (as in Table 11). Accordingly, only a few factors remain statistically significant in this model. The estimated marginal effect of a youth having a job now becomes statistically significant and negative. Higher maternal education and higher family income are still negatively associated with higher depression scores, but both factors become statistically insignificant. It is surprising that in this model, the estimated marginal effect of family poverty status on depression scores becomes negative, which is counter-intuitive.

Table 13 presents the results for the instrumental variable conditional quantile estimation with individual fixed effects based on the jittered sample. The point estimates of the marginal effects are based on 500 jittering replications. The

reported standard errors are based on 499 bootstrap replications. Consistent with the conditional quantile estimation with fixed effects but without instrumenting (as in Table 11), the estimated state dependence parameter is statistically insignificant across all the quantiles. The magnitude of the estimates is much smaller based on the jittered sample. This is consistent with the above conclusion that the state dependence of youth depression is rather mobile in nature at all parts of the distribution when unobserved individual heterogeneity is taken into account. Again, because we have fewer time periods to estimate the model, only a few factors remain statistically significant in this model. The patterns with regard to the effect of the other variables are similar with those observed in the IV estimator without the jittering process (as in Table 12).

2.5 Conclusions

Our study examines the roles of family SES, early childhood life-events, unobserved heterogeneity and pure state dependence in explaining the distribution of depression among adolescents and young adults. We employ a conditional quantile regression framework to address this question and to explore potential heterogeneity in the effects of these factors across different quantiles of the depression score. This is important because these factors of interest may not only

affect the location of the conditional distribution of youth depression, but also affect the scale or other aspect of the distributional shape. If the underlying mechanism that links these factors with youth depression does differ at different parts of the depression distribution, using a conditional mean estimation will neglect this aspect and provide quite different policy implications. Using the US data on the children of the NLSY79 cohort, we first estimated a set of static conditional mean models. The results are in line with the majority of the literature, which highlights the important roles of gender, race, birth order, maternal drinking and smoking behaviour during pregnancy, and a set of family SES factors including maternal education and family poverty status. However, the pooled conditional mean estimation model results suggest that there is no statistically significant effect of stressful life-events including the incidence of family problem and emotional trauma during childhood. This contradicts the studies that document the adverse effect of these events. We then estimated a set of static conditional quantile regression models. Our results reveal the asymmetry of the link between stressful life events and youth depression. The pooled conditional quantile estimation model results show that the stressful life events appears statistically significant for some of the quantiles but not for the others. The asymmetric behaviour of the links is masked by the conditional mean estimation.

This might explain why some studies observe the adverse effect of stressful life events on youth depression while others do not. Furthermore, our conditional quantile regression results provide us with more insights about the differential effects of various factors at different parts of the depression distribution. For example, different types of life-events have different roles across different quantiles of the depression score: the incidence of family problems (family member loss or parental divorce) during childhood plays a more important role at the higher ends of the depression distribution, while the incidence of emotional trauma during childhood plays a more important role at the lower ends. Moreover, the family SES-youth depression gradient observed in the conditional mean estimation models varies substantially across different quantiles of the depression distribution. Specifically, maternal education, maternal unemployment duration and family poverty status are more important at the higher ends of the depression distribution, and are statistically insignificant at the lowest quantile of the depression distribution. This highlights the importance of devoting resources to the individuals with the most severe levels of depression and employing policies that aim to improve these specific family SES conditions for them.

Our study also explicitly models the dynamics of depression during adolescence to early adulthood. We estimate the persistence level of youth

depression and examine the inter-temporal roles of family SES, early childhood life-events, and unobserved heterogeneity in explaining the dynamics of youth depression. A methodological contribution of our study is that in addition to standard dynamic quantile regression models, we employ a newly-developed instrumental variable quantile regression for dynamic panel with fixed-effects model to address this research question. The dynamic conditional quantile regression models revealed the importance of taking into account individual heterogeneity when examining the dynamics of youth depression. The results from the pooled model indicate that there is a strong positive state dependence of youth depression across all quantiles of the CES-D distribution. Also, the magnitude of the state dependence estimate is larger at higher ends of the depression distribution, indicating a higher persistence level for the individuals who have more severe depressive symptoms. In other words, the individuals at higher ends of the CES-D distribution are less likely to recover from depression. This raises our concern towards individuals who experience more severe depression symptoms. However, after taking into account the individual fixed effects, the persistence level of youth depression becomes very close to zero across all the quantiles according to our estimates based on the instrumental variable with fixed effects model. This delivers a different message about the

dynamics of depression during adolescence to early adulthood: the pure state dependence of youth depression is very low and the observed positive association between previous depression and current depression is mainly due to unobserved individual heterogeneity. Furthermore, the estimates from the quantile regression models show the differential inter-temporal effects of stressful life events and family SES factors across the different quantiles of depression distribution. The incidence of family problems or family member loss, and the incidence of emotional trauma during childhood are more important for the individuals who have relatively lower severity level of depression. The family SES-youth depression gradient is steeper at the higher ends of the depression distribution.

By estimating the static conditional mean and conditional quantile models, we examined the relative contributions of a set of important factors in explaining the variability of youth depression. Our results provide a more complete view of the important roles of stressful life events and family SES factors, and a potential explanation for the differences in the existing evidence. Our dynamic models attempt to examine the state dependence of youth depression and the inter-temporal roles of the same set of factors. The IV estimation with fixed effects model could provide us with a bias-corrected estimate of the pure state dependence parameter taking account of the unobserved heterogeneity. The

results indicate that youth depression is rather mobile or transitory in nature. Unfortunately, the short panel of the data led to considerable imprecision for the estimates obtained via the IV quantile regression with fixed effects model. In future studies, the results can be improved if additional waves of data become available for the estimation of dynamic IV models.

References

Anderson, T. W., and C. Hsiao (1981) 'Estimation of Dynamic Models with Error Components' *Journal of the American Statistical Association*, 76, 598–606.

Anderson, T. W., and C. Hsiao (1982) 'Formulation and Estimation of Dynamic Models Using Panel Data' *Journal of Econometrics*, 18, 47–82.

Arellano, M., and S. Bond (1991) 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations' *Review of Economic Studies*, 58, 277–279.

Asarnow, Joan Rosenbaum1; McKowen, James2; Jaycox, Lisa H.3 (2009) 'Improving care for depression: Integrating evidence-based depression treatment within primary care services.' *Treatments for adolescent depression: Theory and practice.* (pp. 159-174). New York, NY, US: Oxford University Press. xii, 328 pp.

Berndt ER, Koran LM, Finkelstein SN, Gelenberg AJ, Kornstein SG, Miller IM, Thase ME, Trapp GA, Keller MB. (2000) 'Lost Human Capital From Early-Onset Chronic Depression' *American Journal of Psychiatry* 157:940-947

Bruce ML, Takeuchi DT, Leaf PJ. (1991) 'Poverty and psychiatric status: Longitudinal evidence from the New Haven Epidemiologic Catchment Area study' *Arch Gen Psychiatry* 48: 470–474.

Cameron A. Colin, and Trivedi Pravin K (2009) 'Microeconometrics using STATA,' STATA Press 2009

Case, A., Fertig, A and Paxson, C. (2005) 'The lasting impact of childhood health and circumstance' *Journal of Health Economics* 24: 365-389

Case, A., Lubotsky, D. and Paxson, C. (2002) 'Economic Status and Health in Childhood: The Origins of the Gradient' *American Economic Review* Vol. 92, 5: 1308-1334

CASPI, A., WRIGHT B., MOFFITT T., SILVA P. A. (1998) 'Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood' *American sociological review*, vol. 63, no3, pp. 424-451 (4 p.1/4)

Chang, Weining . (2009). 'Epidemiology, comorbidity, and course of adolescent depression.' *Treatments for adolescent depression: Theory and practice.* (pp. 3-25). New York, NY, US: Oxford University Press. xii, 328 pp.

Chernozhukov, V., and C. Hansen (2005) 'An IV Model of Quantile Treatment Effects,' *Econometrica*, 73, 245–261.

Chernozhukov, V., and C. Hansen (2006) 'Instrumental Quantile Regression Inference for Structural and Treatment Effects Models' *Journal of Econometrics*, 132, 491–525.

Chernozhukov, V., and C. Hansen (2008) 'Instrumental Variable Quantile Regression: A Robust Inference Approach' *Journal of Econometrics*, 142, 379–398.

Colman, Ian; Wadsworth, Michael E. J.; Croudace, Tim J.; Jones, Peter B.(2007) 'Forty-year psychiatric outcomes following assessment for internalizing disorder in adolescence.' *The American Journal of Psychiatry*, Vol 164(1), pp. 126-133

Contoyannis, P., Jones, A.M., and Rice, N. (2004a) 'Simulation-based inference in dynamic panel ordered probit models: an application to health' *Empirical Economics*, vol. 29: 49-77

Contoyannis, P., Jones, A.M., and Rice, N. (2004b) 'The dynamics of health in the British Household Panel Survey' *Journal of Applied Econometrics*, vol. 19: 473-503

Costello EJ, Compton SN, Keeler G, Angold A. (2003) 'Relationships between poverty and psychopathology: a natural experiment.' *JAMA*. 290(15):2023-9.

Currie, J. (2009) 'Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development' *Journal of Economic Literature,* 47:1,87-122

Currie, J and Stabile, M. (2003) 'Socio-economic status and child health: Why is the relationship stronger for older children?' *American Economic Review* Vol. 93, 5: 1813-1823

Cunha, F., Heckman, J. (2007). 'The Technology of Skill Formation,' *American Economic Review*, vol. 97(2), pages 31-47, May

Dohrenwend BP, Levav I, Shrout PE et al. (1992) 'Socioeconomic status and psychiatric disorders: the causation-selection issue' *Science* 255:946–952.

Essau, Cecilia A. (2009) 'Treatments for adolescent depression: Theory and practice.' (Ed)1, Book Chapter 1, New York, NY, US: Oxford University Press.

Ettner SL. (1996) 'New evidence on the relationship between income and health' *Journal of Health Economics*;15: 67–85.

Fergusson, D.M. & Woodward, L. J. (2002) 'Mental health, educational and social role outcomes of adolescents with depression' *Archives of General Psychiatry*, 59, pp. 225-231.

Fergusson, David M.1; Boden, Joseph M.1; Horwood, L. John (2007) 'Recurrence of major depression in adolescence and early adulthood, and later mental health, educational and economic outcomes' *British Journal of Psychiatry*. Vol 191(2), pp. 335-342

Galvao A. F. (2011) 'Quantile Regression for Dynamic Panel Data with Fixed Effects,' *Journal of Econometrics*, forthcoming.

Galvao A. F. and Montes-Rojas G. V. (2009) 'A Monte Carlo study on bootstrap inference for penalized quantile regression panel data,' Unpublished manuscript, 2009

Gregg, P., Machin, S., (2000) 'Child Development and Success or Failure in the Youth Labor Market' NBER Chapters in: Youth Employment and Joblessness in Advanced Countries, pages 247-288

Goodman, E., Slap, G., and B. Huang, (2003) 'The Public Health Impact of Socioeconomic Status on Adolescent Depression and Obesity' *Am J Public Health*. 2003 November; 93(11): 1844–1850

Graetz B. (1993) 'Health consequences of employment and unemployment: longitudinal evidence for young men and women' *Social Science Medicine*, 36: 715–724.

Greden J. (2001) 'The burden of recurrent depression: causes consequences, and future prospects' *Journal of Clinical Psychiatry* 62(Suppl. 22): 5–9.

Hamilton VH, Merrigan P, Dufresne E.(1997) 'Down and out: estimating the relationship between mental health and unemployment' *Health Economics*, 6: 397–407.

Heckman, J. (2007) 'The economics, technology, and neuroscience of human capability formation' *Proceedings of the National Academy of Sciences of the United States of America,* 104(33): 13250-55

Kato K., Galvao A. F. and Montes-Rojas G. V. (2010) 'Asymptotics and bootstrap inference for panel quantile regression models with fixed effects,' Unpublished manuscript, 2010

Kessler, R., Foster, C., Saunders, W. and Stang, P. (1995) 'Social consequences of psychiatric disorders, I: Educational attainment' *Am J Psychiatry*, 152:1026-1032

Koenker, R., and G. W. Bassett (1978) 'Regression Quantiles' *Econometrica*, 46, 33–49.

Koenker, R., (2004) 'Quantile regression for longitudinal data,' *Journal of Multivariate Analysis*, Elsevier, vol. 91(1), pages 74-89, October.

Lee, CM., and Eden, VA (2009) 'Family-based approaches to the prevention of depression in at-risk children and youth' *Book chapter of 'Treatments for adolescent depression: Theory and practice'*. (pp. 177-214), edited by Essau, Cecilia A. New York, NY, US: Oxford University Press.

Lewinsohn PM, Rhode P, Klein DN, Seeley JR. (1999) 'Natural course of adolescent major depressive disorder, I: continuity into young adulthood.' *J Am Acad Child Adolesc Psychiatry*;38:56-63.

Machado, J.A.F., and Santos Silva, J. M. C., (2005) 'Quantiles for Counts,' *Journal of the American Statistical Association*, American Statistical Association, vol. 100, pages 1226-1237, December.

McLeod J., Kaiser, K., (2004) 'Childhood emotional and behavioral problems and educational attainment' *American sociological review*, vol. 69, no5, pp. 636-658

MIECH R., CASPI A., MOFFITT T. E., WRIGHT B, SILVA P. A. (1999) 'Low socioeconomic status and mental disorders : A longitudinal study of selection and causation during young adulthood' *American journal of sociology*, vol. 104, no4, pp. 1096-1131 (5 p.1/4)

Murphy JM, Olivier DC, Monson RR, Sobol AM, Federman EB, Leighton AH. (1991) 'Depression and anxiety in relation to social status: A prospective epidemiologic study' *Arch Gen Psychiatry*; 48: 223–229.

Radloff, Lenore S. (1977) 'The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.' *Applied Psychological Measurement;* 1: 385-401.

Repetto, P., Caldwell, C., and Zimmerman, M. (2003) 'Trajectories of depressive symptoms among high risk African-American adolescents' *Journal of Adolescence*, Vol. 26, Issue 2: 169-183

Rice DP, Miller LS. (1995) 'The economic burden of affective disorders' *British Journal of Psychiatry. Supplement.* 1995 Apr;(27):34-42.

Robins, J., Rotnitzky, A. and Scharfstein, D. (2000) 'Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models' in *Statistical Models in Epidemiology, the Environment, and Clinical Trials* by Halloran, M. and Berry, D(eds). Springer-Verlag New York

RUSHTON, J., FORCIER, M., and SCHECTMAN, R. (2002) 'Epidemiology of Depressive Symptoms in the National Longitudinal Study of Adolescent Health' *Journal of the American Academy of Child & Adolescent Psychiatry*, Vol. 41, Issue 2, Pages 199-205

Saluja G et al. (2004) 'Prevalence of and risk factors for depressive symptoms among young adolescents' *Archives of Pediatrics and Adolescent Medicine* 158: 760–765.

Shaffer D, Fisher P, Dulcan MK, Davies M, Piacentini J, Schwab-Stone ME, Lahey BB, Bourdon K, Jensen PS, Bird HR, Canino G, Regier DA. (1996) 'The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study' *Journal of the American Academy of Child and Adolescent Psychiatry*. 1996 Jul;35(7):865-77.

Waschbusch, D. A. , Sellers, D. P., LeBlanc, M. M., & Kelley, M. L. (2003). Helpless attributions and depression in adolescents: The roles of anxiety, event valence, and demographics. *Journal of Adolescence* , *26* , 169-183.

Weissman MM, Wolk S, Goldstein RB, et al. (1999) 'Depressed adolescents grown up' *JAMA*.;281:1707-1713.

Zimmerman,F. and Katon, W. (2005) 'Socioeconomic status, depression disparities, and financial strain: what lies behind the income-depression relationship?' *Health Economics* 14: 1197–1215

Table 2.1: Descriptive statistics of variables used for estimation (N= 3,543 individuals)

Variables	Mean	Std. dev.	Median
Youth CES-D depression score	4.576	3.679	4
Youth CES-D=0	11.20%		
Youth age	19.071	2.933	19
Youth Sex (Male)	49.03%		
Youth race: Hispanic	23.59%		
Youth race: Black	36.14%		
Youth race: Non-Hispanic, non-black	40.28%		
Birth order of youth: First	42.91%		
Birth order of youth: Second	34.02%		
Birth order of youth: Third	15.89%		
Birth order of youth: Fourth and above	7.18%		
Youth live in urban area (0-1)	0.777	0.417	1
Youth live in SMSA (0-1)	0.869	0.337	1
Youth has a CPS job[28] (0-1)	0.720	0.449	1
Youth emotional problem in last year (0-1)	0.069	0.253	0
Youth prescription drug for behavior problem (0-1)	0.037	0.189	0
Incidence of child psychiatrist visits (in all survey years)for:			
Emotional trauma, molestation, abuse	0.021	0.183	0
Loss of parents/siblings, divorce	0.056	0.270	0
Age of mother at birth of child	23.675	3.565	24
Mother drinking alcohol during pregnancy (0-1)	0.421	0.494	0
Mother smoking during pregnancy (0-1)	0.312	0.464	0
Highest grade completed by mother date	12.544	2.560	12
Maternal # of weeks unemployed in past calendar year	2.146	8.316	0
Maternal # of weeks unemployed since last interview	4.115	16.770	0
Total annual parental income (in mother's family)	56,309.81	59,576.68	42,000
Poverty status of family in past calendar year (0-1)	0.224	0.417	0.224

Note: only the variables used for model estimation are included in the table.

[28]A CPS job is a job type within the classification used in the Current Population Survey (CPS).

Table 2.2: Transition matrix for the CES-D score over all waves (5 categories)

	0	1-3	4-6	7-11	12-21	Total
0	26.47	**43.08**	20.16	8.64	1.66	100
1-3	13.51	**43.54**	28.56	12.07	2.33	100
4-6	7.41	33.48	**34.79**	20.97	3.34	100
7-11	4.35	20.96	**32.50**	30.65	11.54	100
12-21	3.93	16.16	20.52	**35.15**	24.24	100
Total	11.00	34.87	29.62	18.98	5.53	100

Note: the cells represent the unconditional transition probabilities in percentages. The bold cell shows the biggest cell in each row.

Table 2.3: Static conditional mean estimation for CES-D score—Linear model

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled linear model		Linear model, random-effects specification		Linear model, fixed-effects specification	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
Youth Gender: male	-0.8111***	0.0993	-0.8327***	0.0992		
Race: black	0.4271***	0.1394	0.4704***	0.1380		
Race: non-Hispanic & non-black	0.0626	0.1379	0.0540	0.1405		
Birth order	0.1978***	0.0555	0.1972***	0.0547		
Emotional problem consultation last year	1.9760***	0.2143	1.4452***	0.1600	0.5903***	0.1895
Drug use for behavior problem last year	1.6478***	0.2865	1.2840***	0.2187	0.4596	0.2808
Youth has a CPS job	0.3899***	0.0957	0.3981***	0.0874	0.3915***	0.1039
Incidence of family problem during childhood	0.2960	0.2047	0.3368*	0.1859		
Incidence of emotional trauma during childhood	0.4788	0.3174	0.5495**	0.2680		
Age of mother at birth of child	-0.0529***	0.0165	-0.0512***	0.0160		
Mother drinking during pregnancy	0.1874*	0.1100	0.1800*	0.1065		
Mother smoking during pregnancy	0.2987**	0.1202	0.3572***	0.1143		
Youth living in urban	0.0414	0.1179	0.0279	0.1106	0.0705	0.1516
Youth living in SMSA	0.0488	0.1462	0.0992	0.1375	0.2231	0.1996
Maternal highest grade completed	-0.0408*	0.0235	-0.0366*	0.0219	-0.0671	0.0563
Maternal # of weeks unemployed last year	0.0013	0.0095	0.0039	0.0074	0.0087	0.0089
Maternal # of weeks unemployed since last interview	0.0024	0.0050	0.0004	0.0037	-0.0035	0.0043
Maternal total family income	-7.13E-07	8.90E-07	-6.13E-07	8.30E-07	-2.54E-07	1.19E-06
Maternal family poverty status	0.2562*	0.1324	0.2141*	0.1168	0.0561	0.1639
Maternal family low-income level	-1.31E-05	1.01E-05	-1.21E-05	8.97E-06	3.71E-06	1.54E-05
Constant	5.6418***	0.4790	5.5241***	0.4498	4.7604***	0.7740
sigma_u			1.9904		2.9511	
sigma_e			2.9218		2.9218	
ICC (rho)			0.3170		0.5050	

1. The reported standard errors are robust to cluster effects for the pooled specification.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
3. ICC is the intra-class correlation coefficient, (σ_u^2/ (1+ σ_u^2)).
4. The time-invariant regressors are automatically dropped from the fixed-effects model.

Table 2.4: Static conditional mean estimation for CES-D score—Poisson model

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled model		Poisson model, random-effects specification		Poisson model, fixed-effects specification	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
Youth Gender: male	-0.8015***	0.0978	-0.8071***	0.0996		
Race: black	0.4208***	0.1416	0.4902***	0.1357		
Race: non-Hispanic & non-black	0.0657	0.1412	0.0463	0.1466		
Birth order	0.1872***	0.0515	0.1909***	0.0499		
Emotional problem consultation last year	1.8308***	0.2056	0.8579***	0.1613	0.0979**	0.0397
Drug use for behavior problem last year	1.4158***	0.2604	0.7477***	0.2306	0.0680	0.0606
Youth has a CPS job	0.3901***	0.0933	0.4091***	0.0917	0.0850***	0.0283
Incidence of family problem during childhood	0.2624	0.1713	0.3590**	0.1767		
Incidence of emotional trauma during childhood	0.3402	0.2159	0.4918**	0.2344		
Age of mother at birth of child	-0.0511***	0.0161	-0.0517***	0.0166		
Mother drinking during pregnancy	0.1789*	0.1081	0.1637	0.1086		
Mother smoking during pregnancy	0.2849**	0.1160	0.3884***	0.1146		
Youth living in urban	0.0351	0.1178	0.0313	0.1197	0.0128	0.0321
Youth living in SMSA	0.0477	0.1440	0.1370	0.1328	0.0500	0.0428
Maternal highest grade completed	-0.0423*	0.0235	-0.0394*	0.0230	-0.0149	0.0097
Maternal # of weeks unemployed last year	0.0012	0.0076	0.0053	0.0072	0.0015	0.0017
Maternal # of weeks unemployed since last interview	0.0019	0.0037	-0.0011	0.0033	-0.0005	0.0008
Maternal total family income	-8.81E-07	0.00E+00	-6.07E-07	0.00E+00	-5.00E-08	0.00E+00
Maternal family poverty status	0.2280*	0.1260	0.1514	0.1243	0.0135	0.0364
Maternal family low-income level	-1.30E-05	1.00E-05	-8.20E-06	1.00E-05	9.67E-07	0.00E+00

1. For the pooled specification, the reported standard errors are robust to cluster effects; for the random-effects and fixed-effects models, the reported standard errors are based on bootstrapping for 499 replications.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
3. The time-invariant regressors are automatically dropped from the fixed-effects model.

Table 2.5: Static conditional quantile estimation without individual-specific effects: No jittering process (CES-D score as a continuous variable)

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
Youth Gender: male	-0.3182***	0.0653	-0.6506***	0.0851	-1.2216***	0.1348
Race: black	0.3396***	0.0915	0.4071***	0.1190	0.3005	0.1890
Race: non-Hispanic & non-black	0.0220	0.0932	0.1411	0.1219	0.0236	0.1935
Birth order	0.1582***	0.0361	0.1190**	0.0471	0.2972***	0.0747
Emotional problem consultation last year	1.2602***	0.1379	1.6442***	0.1784	2.4537***	0.2826
Drug use for behavior problem last year	1.1644***	0.1802	1.6988***	0.2347	2.4194***	0.3725
Youth has a CPS job	0.3788***	0.0746	0.2974***	0.0975	0.5237***	0.1544
Incidence of family problem during childhood	0.0426	0.1209	0.2788*	0.1600	0.5207**	0.2520
Incidence of emotional trauma during childhood	0.4406***	0.1692	0.4987**	0.2260	0.3139	0.3571
Age of mother at birth of child	-0.0245**	0.0108	-0.0487***	0.0141	-0.1028***	0.0222
Mother drinking during pregnancy	0.0619	0.0703	0.2079**	0.0912	0.4028***	0.1455
Mother smoking during pregnancy	0.2200***	0.0759	0.2136**	0.0983	0.4004**	0.1564
Youth living in urban	-0.0373	0.0882	0.0585	0.1148	0.2459	0.1828
Youth living in SMSA	0.0520	0.1077	-0.0239	0.1403	-0.1712	0.2223
Maternal highest grade completed	-0.0184	0.0150	-0.0375*	0.0199	-0.0414	0.0317
Maternal # of weeks unemployed last year	0.0003	0.0052	0.0042	0.0079	0.0121	0.0134
Maternal # of weeks unemployed since last interview	0.0008	0.0022	0.0024	0.0038	-0.0038	0.0067
Maternal total family income	-7.98E-07	6.16E-07	-7.11E-07	8.35E-07	-1.61E-06	1.38E-06
Maternal family poverty status	0.0561	0.0918	0.2208*	0.1199	0.2150	0.1901
Maternal family low-income level	-5.91E-06	6.32E-06	-2.38E-05***	8.59E-06	-3.15E-05**	1.41E-05
Constant	2.1959***	0.3138	5.1982***	0.4117	8.9224***	0.6625

1. The reported standard errors are robust to cluster effects for the pooled specification.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.

Table 2.6: Static conditional quantile estimation without individual-specific effects: With jittering process (CES-D score as a discrete variable)

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
Youth Gender: male	-0.3720***	0.0753	-0.6722***	0.0807	-1.1764***	0.1260
Race: black	0.3618***	0.1071	0.4055***	0.1157	0.3001*	0.1760
Race: non-Hispanic & non-black	0.0586	0.1137	0.1449	0.1121	0.0213	0.1783
Birth order	0.1585***	0.0322	0.1179**	0.0464	0.2604***	0.0676
Emotional problem consultation last year	1.2638***	0.1592	1.5567***	0.1997	2.2617***	0.2972
Drug use for behavior problem last year	0.9631***	0.2227	1.6202***	0.2440	2.0252***	0.2963
Youth has a CPS job	0.3736***	0.0825	0.3297***	0.0919	0.4916***	0.1314
Incidence of family problem during childhood	0.0493	0.1786	0.2630*	0.1441	0.4805	0.3014
Incidence of emotional trauma during childhood	0.3783**	0.1651	0.3915**	0.1658	0.2368	0.3996
Age of mother at birth of child	-0.0213*	0.0129	-0.0498***	0.0134	-0.1000***	0.0205
Mother drinking during pregnancy	0.0439	0.0810	0.1883**	0.0874	0.3895***	0.1337
Mother smoking during pregnancy	0.1826**	0.0873	0.2354**	0.0963	0.3997***	0.1446
Youth living in urban	-0.0357	0.1074	0.0381	0.1122	0.2494	0.1597
Youth living in SMSA	0.0077	0.1300	-0.0036	0.1329	-0.1626	0.1922
Maternal highest grade completed	-0.0142	0.0161	-0.0337*	0.0178	-0.0581*	0.0317
Maternal # of weeks unemployed last year	0.0020	0.0054	0.0029	0.0055	0.0099*	0.0060
Maternal # of weeks unemployed since last interview	0.0004	0.0018	0.0021	0.0020	-0.0033	0.0024
Maternal total family income	-1.18E-06	9.44E-07	-1.09E-06	9.55E-07	-1.52E-06	1.23E-06
Maternal family poverty status	0.0906	0.1043	0.2016*	0.1224	0.1679	0.1624
Maternal family low-income level	-4.48E-06	6.61E-06	-1.96E-05**	8.77E-06	-2.96E-05**	1.22E-05

1. All the estimates are based on 500 jittering replications.
2. The marginal effects are calculated based on the jittered sample.
3. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.

Table 2.7: Dynamic conditional mean estimation for CES-D score—Linear model

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled linear model		Linear model, random-effects specification		Linear model, fixed-effects specification	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	0.3419***	0.0151	0.2351***	0.0127	-0.3702***	0.0175
Youth Gender: male	-0.4876***	0.0871	-0.5914***	0.1010		
Race: black	0.4068***	0.1214	0.4522***	0.1397		
Race: non-Hispanic & non-black	0.1570	0.1217	0.1648	0.1429		
Birth order	0.1221**	0.0479	0.1472***	0.0552		
Emotional problem consultation last year	1.3490***	0.2352	1.3328***	0.1926	0.5496**	0.2384
Drug use for behavior problem last year	0.8686***	0.3082	0.9559***	0.2688	0.6339*	0.3667
Youth has a CPS job	0.1097	0.1162	0.1489	0.1187	0.3674**	0.1511
Incidence of family problem during childhood	0.3036*	0.1596	0.3594*	0.1886		
Incidence of emotional trauma during childhood	0.5318**	0.2613	0.5861**	0.2719		
Age of mother at birth of child	-0.0150	0.0145	-0.0253	0.0164		
Mother drinking during pregnancy	0.1574	0.0961	0.1738	0.1074		
Mother smoking during pregnancy	0.1641	0.1054	0.2224*	0.1160		
Youth living in urban	0.0477	0.1205	0.0645	0.1258	0.2192	0.1827
Youth living in SMSA	0.1005	0.1578	0.1386	0.1730	0.6721**	0.3149
Maternal highest grade completed	-0.0439**	0.0208	-0.0447*	0.0229	-0.0361	0.0646
Maternal # of weeks unemployed last year	0.0092	0.0158	0.0133	0.0153	0.0448**	0.0185
Maternal # of weeks unemployed since last interview	-0.0071	0.0091	-0.0086	0.0089	-0.0210*	0.0109
Maternal total family income	-7.15E-07	7.71E-07	-7.34E-07	8.84E-07	-7.83E-08	1.37E-06
Maternal family poverty status	0.2519**	0.1262	0.2736**	0.1315	0.1887	0.2123
Maternal family low-income level	-7.49E-06	9.01E-06	-6.55E-06	9.39E-06	3.09E-05*	1.83E-05
Constant	3.3863***	0.4437	3.9800***	0.4792	5.0001***	0.9255
sigma_u			1.3643		4.0594	
sigma_e			2.6285		2.6285	
ICC (rho)			0.2122		0.7046	

1. The reported standard errors are robust to cluster effects for the pooled specification.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
3. ICC is the intra-class correlation coefficient, ($\sigma_u^2/ (1+ \sigma_u^2)$).
4. The time-invariant regressors are automatically dropped from the fixed-effects model.

Table 2.8: Dynamic conditional mean estimation for CES-D score—Poisson model

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled model		Poisson model, random-effects specification		Poisson model, fixed-effects specification	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	0.2746***	0.0106	-0.0019	0.0181	-0.0743***	0.0197
Youth Gender: male	-0.4608***	0.0868	-0.8055***	0.1158		
Race: black	0.3947***	0.1225	0.5430***	0.1596		
Race: non-Hispanic & non-black	0.1618	0.1256	0.1609	0.1622		
Birth order	0.1106***	0.0432	0.1935***	0.0575		
Emotional problem consultation last year	1.1171***	0.2100	0.9528***	0.2161	0.1056	0.0694
Drug use for behavior problem last year	0.5518**	0.2476	0.8209***	0.3006	0.1615	0.1071
Youth has a CPS job	0.1291	0.1115	0.2455*	0.1321	0.1210**	0.0555
Incidence of family problem during childhood	0.2406*	0.1267	0.4390**	0.1749		
Incidence of emotional trauma during childhood	0.3599**	0.1710	0.5850**	0.2655		
Age of mother at birth of child	-0.0156	0.0140	-0.0477***	0.0178		
Mother drinking during pregnancy	0.1523*	0.0923	0.2049*	0.1205		
Mother smoking during pregnancy	0.1373	0.0987	0.3708***	0.1379		
Youth living in urban	0.0625	0.1188	0.1197	0.1321	0.0479	0.0487
Youth living in SMSA	0.0578	0.1533	0.2319	0.1738	0.2289**	0.1122
Maternal highest grade completed	-0.0449**	0.0204	-0.0449	0.0284	0.0019	0.0222
Maternal # of weeks unemployed last year	0.0060	0.0141	0.0217	0.0153	0.0117**	0.0059
Maternal # of weeks unemployed since last interview	-0.0054	0.0082	-0.0121	0.0092	-0.0055*	0.0033
Maternal total family income	-8.25E-07	0.0000	-8.95E-07	0.00E+00	-4.86E-08	0.00E+00
Maternal family poverty status	0.2048*	0.1165	0.2687*	0.1441	0.0542	0.0627
Maternal family low-income level	-8.63E-06	1.00E-05	-1.69E-06	1.00E-05	7.97E-06	1.00E-05

1. For the pooled specification, the reported standard errors are robust to cluster effects; for the random-effects and fixed-effects models, the reported standard errors are based on bootstrapping for 499 replications.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
3. The time-invariant regressors are automatically dropped from the fixed-effects model.

Table 2.9: Dynamic conditional quantile estimation without individual-specific effects: No jittering process (CES-D score as a continuous variable)

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	0.2738***	0.0137	0.3624***	0.0118	0.4574***	0.0182
Youth Gender: male	-0.1361	0.0944	-0.3730***	0.0865	-0.8038***	0.1344
Race: black	0.4054***	0.1315	0.4239***	0.1197	0.2996	0.1850
Race: non-Hispanic & non-black	0.1445	0.1330	0.1984	0.1228	0.1294	0.1898
Birth order	0.0878*	0.0514	0.0083	0.0470	0.1866**	0.0734
Emotional problem consultation last year	0.9384***	0.2015	1.1797***	0.1834	1.5569***	0.2867
Drug use for behavior problem last year	0.7001**	0.2775	0.8675***	0.2518	0.7315*	0.3941
Youth has a CPS job	0.3127**	0.1227	0.0615	0.1123	0.1558	0.1743
Incidence of family problem during childhood	0.3993**	0.1780	0.1643	0.1621	0.1625	0.2443
Incidence of emotional trauma during childhood	0.5257**	0.2474	0.5626**	0.2312	0.2182	0.3438
Age of mother at birth of child	-0.0002	0.0154	-0.0024	0.0141	-0.0320	0.0218
Mother drinking during pregnancy	-0.1685*	0.1005	0.0778	0.0919	0.4264***	0.1442
Mother smoking during pregnancy	0.2811***	0.1077	0.1464	0.0993	0.0959	0.1547
Youth living in urban	0.1056	0.1289	-0.0900	0.1164	0.2128	0.1815
Youth living in SMSA	-0.0511	0.1734	-0.0474	0.1567	-0.1494	0.2433
Maternal highest grade completed	-0.0140	0.0225	-0.0519***	0.0200	-0.0647**	0.0306
Maternal # of weeks unemployed last year	-0.0098	0.0156	0.0152	0.0147	-0.0030	0.0250
Maternal # of weeks unemployed since last interview	0.0009	0.0091	-0.0076	0.0086	-0.0010	0.0144
Maternal total family income	-6.48E-07	8.47E-07	-9.79E-07	8.00E-07	-2.20E-06*	1.26E-06
Maternal family poverty status	0.1152	0.1306	0.1535	0.1200	0.1489	0.1852
Maternal family low-income level	5.41E-06	8.99E-06	-1.63E-05*	8.36E-06	-2.14E-05	1.31E-05
Constant	0.3915	0.4583	3.2626***	0.4225	5.6848***	0.6648

1. The reported standard errors are robust to cluster effects for the pooled specification.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.

Table 2.10: Dynamic conditional quantile estimation without individual-specific effects: With jittering process (CES-D score as a discrete variable)

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	0.2091***	0.0129	0.2986***	0.0119	0.3869***	0.0167
Youth Gender: male	-0.1495	0.0993	-0.3716***	0.0918	-0.7518***	0.1344
Race: black	0.4458***	0.1487	0.4968***	0.1243	0.3055	0.2015
Race: non-Hispanic & non-black	0.2136	0.1483	0.2808**	0.1296	0.1034	0.1981
Birth order	0.0871*	0.0502	0.0244	0.0483	0.2026**	0.0812
Emotional problem consultation last year	0.7691***	0.2559	1.0770***	0.2149	1.2845***	0.4128
Drug use for behavior problem last year	0.3921	0.2714	0.5800**	0.2479	0.5835	0.3697
Youth has a CPS job	0.2964**	0.1223	0.1024	0.1119	0.1464	0.1898
Incidence of family problem during childhood	0.2289	0.1617	0.1358	0.1374	0.0764	0.1812
Incidence of emotional trauma during childhood	0.3074	0.2056	0.4599***	0.1621	0.2415	0.6581
Age of mother at birth of child	0.0023	0.0147	-0.0080	0.0148	-0.0399*	0.0242
Mother drinking during pregnancy	-0.1272	0.1033	0.0843	0.1000	0.4588***	0.1385
Mother smoking during pregnancy	0.2064*	0.1116	0.1128	0.1063	0.0750	0.1529
Youth living in urban	0.0920	0.1353	-0.0801	0.1344	0.1929	0.1581
Youth living in SMSA	-0.0470	0.1678	-0.0130	0.1877	-0.2189	0.2296
Maternal highest grade completed	-0.0119	0.0209	-0.0491**	0.0222	-0.0638*	0.0333
Maternal # of weeks unemployed last year	-0.0036	0.0174	0.0079	0.0188	0.0042	0.0291
Maternal # of weeks unemployed since last interview	-0.0024	0.0094	-0.0049	0.0122	-0.0040	0.0159
Maternal total family income	-1.30E-06	8.96E-07	-1.04E-06	7.40E-07	-2.11E-06	1.29E-06
Maternal family poverty status	0.0704	0.1261	0.1850	0.1279	0.1626	0.1747
Maternal family low-income level	3.44E-06	7.53E-06	-1.86E-05**	8.20E-06	-2.50E-05**	1.24E-05

1. All the estimates are based on 500 jittering replications.
2. The marginal effects are calculated based on the jittered sample.
3. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.

Table 2.11: Dynamic conditional quantile estimation with individual fixed effects: No jittering process (CES-D score as a continuous variable) ----- Koenker 2004 method

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	-0.4611***	0.0256	-0.3607***	0.0241	-0.2828***	0.0252
Youth Gender: male						
Race: black						
Race: non-Hispanic & non-black						
Birth order						
Emotional problem consultation last year	0.3223	0.2587	0.3341	0.2532	0.6160*	0.3430
Drug use for behavior problem last year	0.6855**	0.3424	0.4930	0.3588	0.0041	0.4393
Youth has a CPS job	0.2074	0.1528	0.1817	0.1520	0.2543*	0.1502
Incidence of family problem during childhood						
Incidence of emotional trauma during childhood						
Age of mother at birth of child						
Mother drinking during pregnancy						
Mother smoking during pregnancy						
Youth living in urban	0.0369	0.1951	0.1174	0.1822	0.2625	0.1748
Youth living in SMSA	0.0763	0.3289	0.2095	0.3671	0.2170	0.3277
Maternal highest grade completed	-0.0320	0.0634	-0.0187	0.0641	-0.0219	0.0652
Maternal # of weeks unemployed last year	0.0029	0.0189	0.0261	0.0196	0.0551***	0.0203
Maternal # of weeks unemployed since last interview	-0.0040	0.0101	-0.0126	0.0100	-0.0227**	0.0115
Maternal total family income	-1.08E-06	1.46E-06	-2.84E-07	1.49E-06	-2.97E-07	1.49E-06
Maternal family poverty status	0.1269	0.2285	0.2021	0.2196	0.1948	0.2249
Maternal family low-income level	2.05E-05	1.81E-05	1.62E-05	1.70E-05	3.23E-05*	1.86E-05

1. The reported standard errors are based on 499 bootstrapping replications.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
3. The time-invariant regressors are dropped from the fixed-effects model.

Table 2.12: Dynamic conditional quantile estimation: instrumental variable approach with individual fixed effects: No jittering process (CES-D score as a continuous variable) ----- Galvao 2011 method

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	-0.1444	0.1818	-0.0058	0.1596	0.0813	0.1521
Youth Gender: male						
Race: black						
Race: non-Hispanic & non-black						
Birth order						
Emotional problem consultation last year	0.9881	0.6561	0.9914	0.6450	0.9797	0.6417
Drug use for behavior problem last year	-0.0989	1.4006	-0.3526	1.4256	-0.5013	1.3997
Youth has a CPS job	-1.0800**	0.4911	-1.0722**	0.4855	-1.0796**	0.4831
Incidence of family problem during childhood						
Incidence of emotional trauma during childhood						
Age of mother at birth of child						
Mother drinking during pregnancy						
Mother smoking during pregnancy						
Youth living in urban	0.1312	0.5345	0.0998	0.5406	0.0897	0.5437
Youth living in SMSA	-0.7634	0.7966	-0.7868	0.7928	-0.7936	0.7987
Maternal highest grade completed	-0.0086	0.1631	-0.0135	0.1625	-0.0157	0.1629
Maternal # of weeks unemployed last year	-0.0011	0.0478	0.0035	0.0477	0.0065	0.0479
Maternal # of weeks unemployed since last interview	0.0020	0.0268	-0.0018	0.0266	-0.0044	0.0266
Maternal total family income	-2.30E-06	3.45E-06	-2.12E-06	3.48E-06	-1.96E-06	3.46E-06
Maternal family poverty status	-1.2136**	0.5922	-1.2488**	0.5909	-1.2683**	0.5897
Maternal family low-income level	-1.07E-05	3.76E-05	-5.21E-06	3.74E-05	-2.80E-06	3.76E-05
Instrumental variable: CESD lag (t-2)	0.0353	0.0593	0.0009	0.0602	-0.0205	0.0676

1. The reported standard errors are based on 499 bootstrapping replications.
2. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
3. The time-invariant regressors are dropped from the fixed-effects model.

Table 2.13: Dynamic conditional quantile estimation: instrumental variable approach with individual fixed effects: With jittering process (CES-D score as a discrete variable) ----- Galvao 2011 method with jittered sample

	(1)	(2)	(3)	(4)	(5)	(6)
	0.25 Quantile regression		0.50 Quantile regression		0.75 Quantile regression	
	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.	Marg. Eff.	St. Err.
CESDlag (t-1)	-0.0193	0.0432	-0.0140	0.0343	-0.0137	0.0368
Youth Gender: male						
Race: black						
Race: non-Hispanic & non-black						
Birth order						
Emotional problem consultation last year	0.3725	0.2520	0.2951	0.1942	0.3036	0.1964
Drug use for behavior problem last year	-0.1371	0.5386	-0.1179	0.5717	-0.1203	0.6386
Youth has a CPS job	-0.5308*	0.2906	-0.4038*	0.2317	-0.3894	0.2438
Incidence of family problem during childhood						
Incidence of emotional trauma during childhood						
Age of mother at birth of child						
Mother drinking during pregnancy						
Mother smoking during pregnancy						
Youth living in urban	0.1016	0.2243	0.0691	0.1728	0.0624	0.1653
Youth living in SMSA	-0.4018	0.5442	-0.3044	0.3773	-0.3130	0.4109
Maternal highest grade completed	0.0043	0.0812	0.0019	0.0622	0.0001	0.0612
Maternal # of weeks unemployed last year	0.0102	0.0215	0.0077	0.0165	0.0071	0.0167
Maternal # of weeks unemployed since last interview	-0.0053	0.0127	-0.0039	0.0098	-0.0034	0.0099
Maternal total family income	-1.17E-06	1.18E-06	-9.20E-07	9.03E-07	-9.67E-07	9.17E-07
Maternal family poverty status	-0.4497**	0.2086	-0.3409**	0.1606	-0.3381**	0.1642
Maternal family low-income level	-3.59E-06	1.64E-05	-2.72E-06	1.25E-05	-2.74E-06	1.22E-05
Instrumental variable: CESD lag (t-2)	-0.0009	0.0065	-0.0006	0.0048	-0.0002	0.0046

1. The point estimates of the marginal effects are based on 500 jittering replications.
2. The reported standard errors are based on 499 bootstrapping replications.
3. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level, * denotes statistical significance at 10% level.
4. The time-invariant regressors are dropped from the fixed-effects model.

Figure 2.1: Histogram of youth CES-D depression score

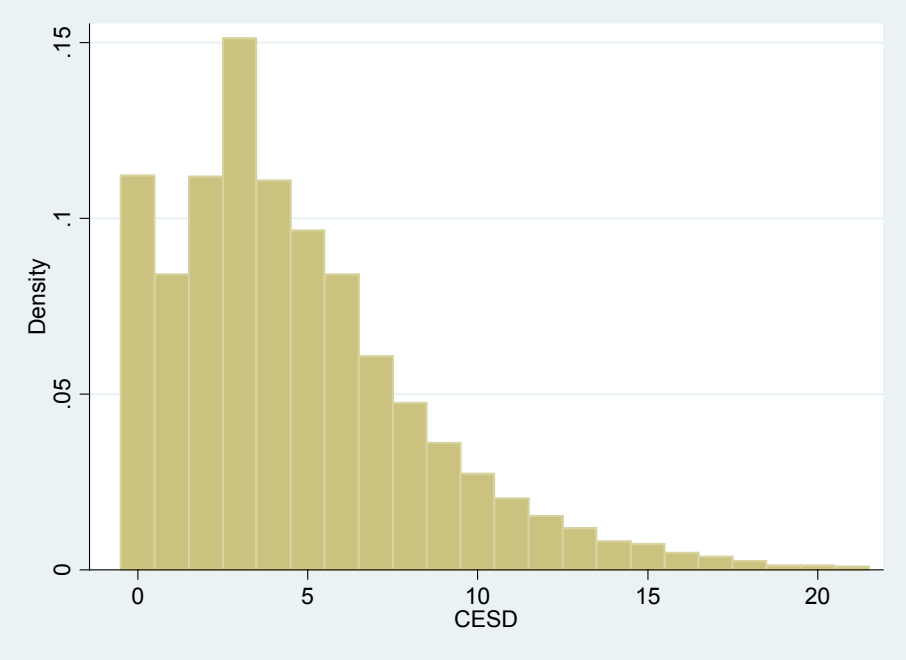


Figure 2.2: Dynamic quantile regression pooled estimates of marginal effects and 95% confidence intervals by quantiles

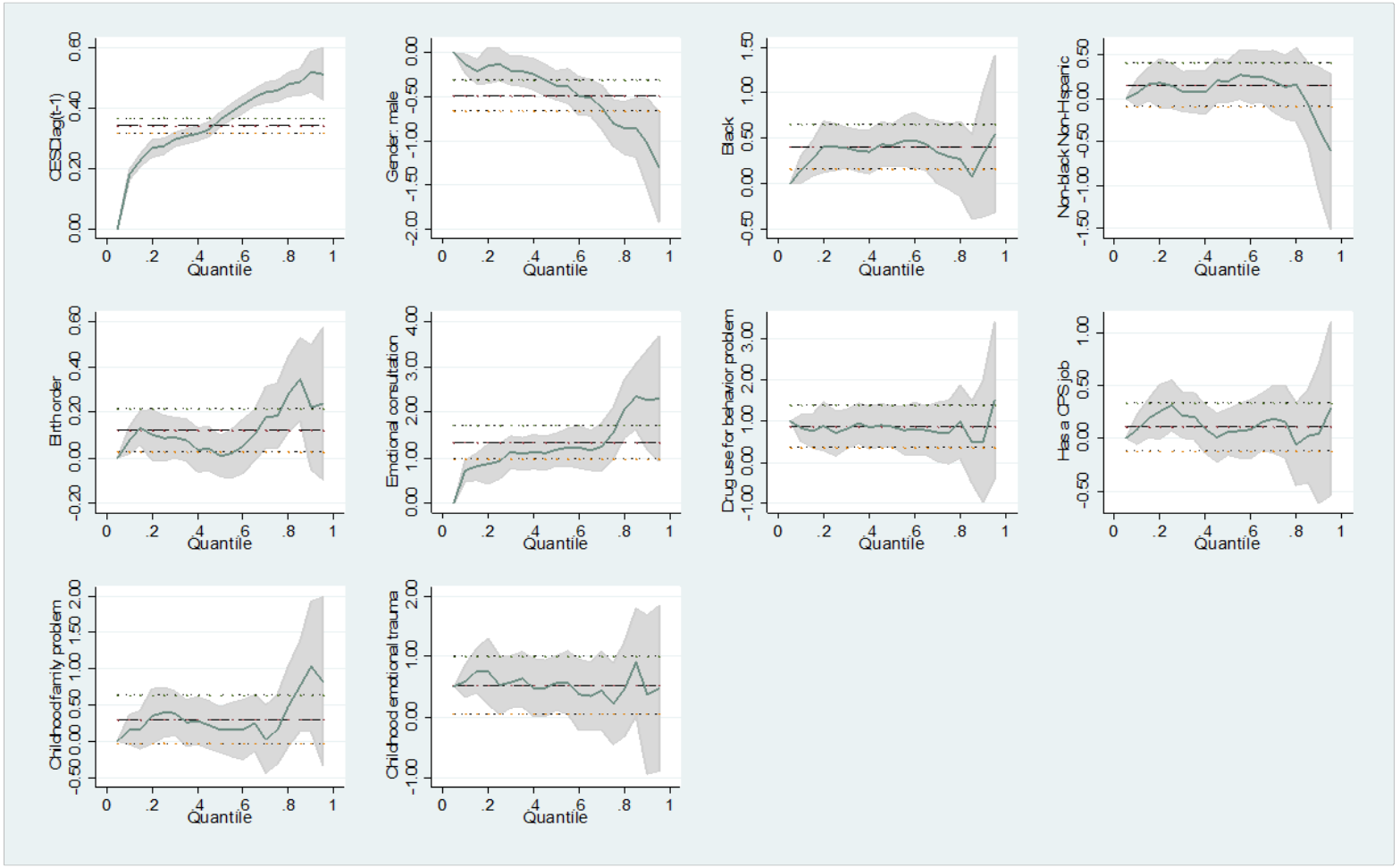
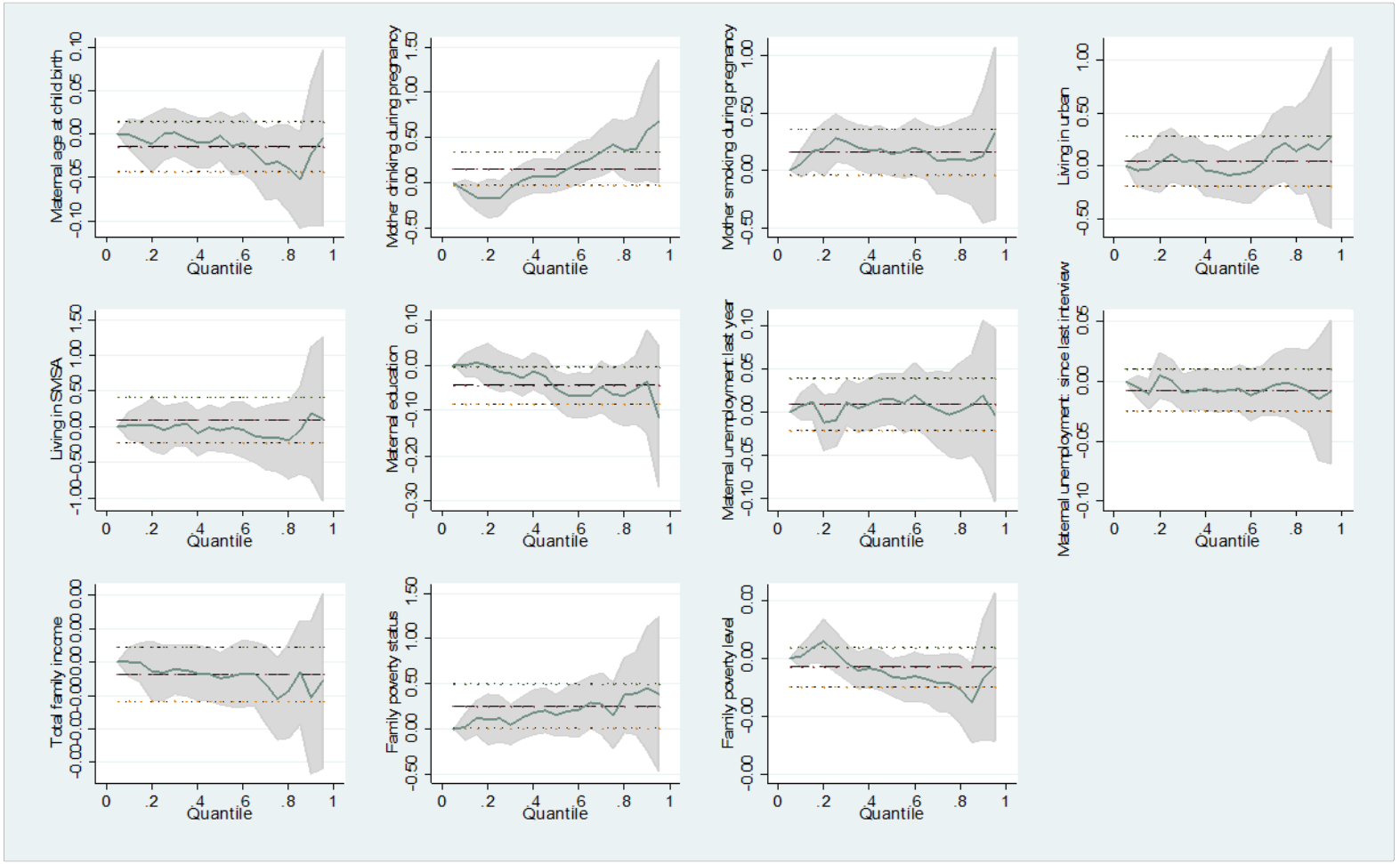


Figure 2.2: Dynamic quantile regression pooled estimates of marginal effects and 95% confidence intervals by quantiles (Continued)



2.A1 Appendix: Variable definitions

Table 2.A.1 Variable definition and corresponding survey questions

Type of variable/ area of interests	**Variable definition**	**Corresponding questions and coding in the survey**
Dependent variables	youth CES-D scale-- main dependent variable; a 7-item scale	CESD - POOR APPETITE; CESD - TROUBLE KEEPING MIND ON TASKS; CESD - DEPRESSED; CESD - EVERYTHING TOOK EXTRA EFFORT; CESD - RESTLESS SLEEP; CESD - SAD; CESD - COULD NOT GET GOING 0 Rarely, None of the time, 1 Day; 1 Some, A little of the time, 1-2 days; 2 Occasionally, Moderate Amt. of the time, 3-4 days; 3 Most, All of the time, 5-7 days
Independent variables: depression at childhood	*History*--child depression scale-- depression at childhood age; a 9-item scale	DEPRESSION - HOW OFTEN CHILD FEELS SAD AND BLUE; DEPRESSION - HOW OFTEN CHILD FEELS NERVOUS, TENSE, OR ON EDGE; DEPRESSION - HOW OFTEN CHILD FEELS HAPPY; DEPRESSION - HOW OFTEN CHILD FEELS BORED; DEPRESSION - HOW OFTEN CHILD FEELS LONELY; DEPRESSION - HOW OFTEN CHILD FEELS TIRED OR WORN OUT; DEPRESSION - HOW OFTEN CHILD FEELS EXCITED ABOUT SOMETHING; DEPRESSION - HOW OFTEN CHILD FEELS TOO BUSY TO GET EVERYTHING; DEPRESSION - HOW OFTEN CHILD FEELS PRESSURED BY MOM OR DAD 1 Often; 2 Sometimes; 3 Hardly ever
Independent variables: youth demographics	Gender-- sex of youth	1=male, 2=female
	Age-- age of youth at assessment date	AGE OF YOUNG ADULT (IN YEARS) AT DATE OF INTERVIEW
	Race of youth	1=Hispanic, 2=Black, 3=non-hispanic non-black
	Birth order of youth	Birth order of child, range 1 to 11
	Youth marital status	CURRENT MARITAL STATUS
Independent variables: home environment/ living area	*History*-- Child living with parents in the household	Child usual residence-- living in a household with parents. 1 IN HOUSEHOLD OF MOTHER 2 WITH FATHER 3 WITH OTHER RELATIVE(S) 4 WITH FOSTER CARE 5 WITH ADOPTIVE PARENT(S) 6 IN LONG TERM CARE INSTITUTION 7 AWAY AT SCHOOL 8 DECEASED 9 PART TIME W/ MOTHER, PART TIME W/ FATHER 10 PART TIME W/ MOTHER, PART TIME W/ OTHER 11 OTHER

	Youth live in residence own unit/ parental household/other	TYPE OF RESIDENCE R LIVES IN-- 11 OWN DWELLING UNIT 19 RESPONDENT IN PARENTS' HOUSEHOLD (BOTH PARENTS PRESENT) 20 RESPONDENT IN MOTHER'S HOUSEHOLD 21 RESPONDENT IN FATHER'S HOUSEHOLD 22 RESPONDENT IN OTHER RELATIVE'S HOUSEHOLD 15 CONVENT, MONASTERY, OTHER RELIGIOUS INSTITUTE 13 OFF-BASE MILITARY FAMILY HOUSING 12 ON-BASE MILITARY FAMILY HOUSING 16 OTHER INDIVIDUAL QUARTERS (SPECIFY)
	Youth live in rural/urban	IS CURRENT RESIDENCE URBAN OR RURAL?
	Youth live in SMSA	IS CURRENT RESIDENCE IN SMSA?
Independent variables: cognitive/non-cognitive abilities	*History—* Child Cognitive abilities—PIAT Math score	PIAT MATH: TOTAL RAW SCORE PIAT MATH: TOTAL STANDARD SCORE PIAT MATH: TOTAL PERCENTILE SCORE (this instrument is administered for the age range of 5-14 years old)
	History— Child Cognitive abilities—PIAT Reading recognition score	PIAT READING RECOGNITION : TOTAL RAW SCORE PIAT READING RECOGNITION : TOTAL STANDARD SCORE PIAT READING RECOGNITION : TOTAL PERCENTILE SCORE (this instrument is administered for the age range of 5-14 years old)
	History— Child Cognitive abilities—PIAT Reading comprehension score	PIAT READING COMPREHENSION: TOTAL RAW SCORE PIAT READING COMPREHENSION: TOTAL STANDARD SCORE PIAT READING COMPREHENSION: TOTAL PERCENTILE SCORE (this instrument is administered for the age range of 5-14 years old)
	History— Child Cognitive abilities—PPVT (PEABODY PICTURE VOCABULARY TEST) score	PEABODY PICTURE VOCABULARY TEST: TOTAL RAW SCORE PEABODY PICTURE VOCABULARY TEST: TOTAL STANDARD SCORE PEABODY PICTURE VOCABULARY TEST: TOTAL PERCENTILE SCORE (this instrument is administered for the age range of 4-5, and again for the range of 10-11 years old)
Independent variables: stressful life events	*History—* Child Emotional disturbance	CHILD HAS SERIOUS EMOTIONAL DISTURBANCE (available for 1986-2000 direct variable, 2002-2008 loops of questions) Any health condition/limitation? What is her health condition/ limitation? One choice is--SERIOUS EMOTIONAL DISTURBANCE
	History— Child PSYCH PROBLEM-EMOTIONAL TRAUMA, MOLESTATION, ABUSE	(Available for 1988-2008) During the past 12 months has the respondent seen a psychiatrist, because of – EMOTIONAL TRAUMA, MOLESTATION, ABUSE
	History— Child PSYCH PROBLEM-FAMILY PROBLEMS OR LOSS	(Available for 1988-2008) During the past 12 months has the respondent seen a psychiatrist, because of – FAMILY PROBLEMS OR LOSS (loss of parents/siblings, divorce)
Independent variables: health care utilization	*History—* Child psychologist consultation in last year	During the past 12 months has the respondent seen a psychiatrist, psychologist, or counselor about any behavioral, emotional, or mental problem? 1 YES, 0 NO
	Youth emotional problem consultation in last year	During the last 12 months, have you received any help for an emotional, behavioral, or family problem?

	Youth prescription drug use for behavior problem in last year	Do you regularly take any medicine or prescription drugs to help control your activity level or behavior?
Independent variables: Parental level-SES	Maternal education-- Highest grade completed	(Included in the Child Interview) HIGHEST GRADE COMPLETED BY MOTHER AS OF DATE OF INTERVIEW
	Maternal education-- mother currently enrolled in school	(Included in the Child Interview) MOTHER CURRENTLY ATTENDING OR ENROLLED IN REGULAR SCHOOL? 1 YES, 0 NO
	Maternal employment status	**Constructed from the NLSY79 main survey** NUMBER OF WEEKS UNEMPLOYED SINCE LAST INTERVIEW (1994-2008)
	Maternal employment status	**Constructed from the NLSY79 main survey** NUMBER OF WEEKS UNEMPLOYED IN PAST CALENDAR YEAR (1994-2008)
	Maternal employment status	**Constructed from the NLSY79 main survey** EMPLOYMENT STATUS (only available for 1994-1998 and 2006)
	Paternal education	**Included in the Youth Interview** What is the HIGHEST GRADE ever COMPLETED BY your FATHER? (1994-2008)
	Paternal employment status	**Included in the Youth Interview** DID FATHER WORK FOR PAY ALL OF last year, PART, OR NOT AT ALL?
	Paternal employment status	**Included in the Youth Interview** OCCUPATION OF LONGEST JOB OF FATHER IN 1993 (1970 CENSUS 3 DIGIT)
	family income in the family (of mother's)	**Constructed from the NLSY79 main survey** TOTAL NET FAMILY INCOME IN PAST CALENDAR YEAR--income from all sources from the respondent and the spouse
	financial difficulties in the family (of mother's)	**Constructed from the NLSY79 main survey** Family poverty status in past calendar year
	low-income level in the family (of mother's)	**Constructed from the NLSY79 main survey** Family poverty level in past calendar year
Independent variables: Youth-SES (NOTE: these variables are relevant when youth is "emancipated")	Status of youth—emancipated or not (question starting from 2000)	IS R UNEMANCIPATED, AS DEFINED IN Q15-1B? ([flag indicating that R is under 18, living with at least one parent, not married, not living with a partner, and has no children]=1)
	Youth education	YEAR OF SCHOOL/GRADE R IS CURRENTLY ENROLLED IN
	Youth education	HAS R EVER REPEATED A GRADE?
	Youth education	HAS R EVER SKIPPED AHEAD A GRADE?
	Youth education	HIGHEST GRADE OF REGULAR SCHOOL R HAS COMPLETED
	Youth employment status	R HAS A CPS JOB?
	Youth last job	EVER WORKED PART-TIME OR FULL-TIME AT JOB LASTING TWO CONSECUTIVE WEEKS OR MORE
	Youth last job—industry	BUSINESS OR INDUSTRY AT LAST JOB LASTING TWO WEEKS OR MORE (2000 CENSUS 3 DIGIT)
	Youth last job—occupation categories	OCCUPATION AT LAST JOB LASTING TWO WEEKS OR MORE (2000 CENSUS 3 DIGIT)
	Youth income—own income	TOTAL INCOME FROM MILITARY SERVICE in the last year TOTAL INCOME FROM WAGES AND SALARY in the last year TOTAL INCOME FROM FARM OR BUSINESS in the last year -- Need to sum up

	Youth income—income from spouse/partner (if relevant)	TOTAL INCOME of spouse/partner FROM MILITARY SERVICE in the last year TOTAL INCOME of spouse/partner FROM WAGES AND SALARY in the last year TOTAL INCOME of spouse/partner FROM FARM OR BUSINESS in the last year -- Need to sum up
	Youth income—total income of youth and spouse from other income sources of social security	THE TOTAL AMOUNT R/SPOUSE/PARTNER RECEIVED FROM THESE OTHER BENEFITS in the last year What was the total amount of these (other) veterans benefits, worker's compensation, disability payments, or payments from Social Security [R or R's spouse or partner] received during
Independent variables: Medical/biological factors	age of mother at birth of child	Age of Mother at birth of child, 8 categories
	mother drinking alcohol during 1 year before birth	MOTHER DRINK ALCOHOL DURING 12 MONTHS BEFORE BIRTH OF CHILD? 1 YES, 0 NO
	frequency of alcohol use during pregnancy	FREQUENCY OF ALCOHOL USE BY MOTHER DURING PREGNANCY -- 0 NEVER, 1 LESS THAN ONCE A MONTH, 2 ABOUT ONCE A MONTH, 3 3 OR 4 DAYS A MONTH, 4 1 OR 2 DAYS A WEEK, 5 3 OR 4 DAYS A WEEK, 6 NEARLY EVERY DAY, 7 EVERY DAY
	mother smoking during 1 year before birth	MOTHER SMOKE DURING 12 MONTHS BEFORE BIRTH OF CHILD?
	# of cigarettes smoked during pregnancy	# OF CIGARETTES SMOKED BY MOTHER DURING PREGNANCY-- 0 DID NOT SMOKE, 1 LESS THAN 1 PACK A DAY, 2 1 OR MORE BUT LESS THAN 2, 3 2 OR MORE PACKS A DAY
	Substance (marijuana/hashish/cocaine) use during 1 year before birth	MOTHER USE MARIJUANA/HASHISH DURING 12 MONTHS BEFORE BIRTH OF CHILD? Or MOTHER USE COCAINE DURING 12 MONTHS BEFORE BIRTH OF CHILD?
	frequency of Substance (marijuana/hashish/cocaine) use during pregnancy	FREQUENCY OF MARIJUANA/HASHISH USE BY MOTHER DURING PREGNANCY Or FREQUENCY OF COCAINE USE BY MOTHER DURING PREGNANCY
Instrumental variables: SES of grandparents	Mother/stepmother (of mother's)—work status	(only in 1979 survey of NLSY79 main survey) DID MOTHER/STEPMOTHER WORK FOR PAY ALL OF 1978, PART, OR NOT AT ALL?
	Mother/stepmother (of mother's)—occupation class	OCCUPATION OF LONGEST JOB IN 1978, R'S MOTHER/STEPMOTHER (CENSUS 3 DIGIT)
	Mother/stepmother (of mother's)—work status	DID MOTHER/STEPMOTHER WORK > 35 HOURS PER WEEK IN 1978?
	Mother/stepmother (of mother's)—education	HIGHEST GRADE COMPLETED BY R'S MOTHER
	Father/stepfather (of mother's)—work status	DID FATHER/STEPFATHER WORK FOR PAY ALL OF 1978, PART, OR NOT AT ALL?
	Father/stepfather (of mother's)—occupation class	OCCUPATION OF LONGEST JOB IN 1978, R'S FATHER/STEPFATHER (CENSUS 3 DIGIT)
	Father/stepfather (of mother's)—work status	DID FATHER/STEPFATHER WORK > 35 HOURS PER WEEK IN 1978?
	Father/stepfather (of mother's)—education	HIGHEST GRADE COMPLETED BY R'S FATHER

	Adult female figure (in the household of mother's)—work for pay	DID ADULT MALE PRESENT IN HOUSEHOLD AT AGE 14 WORK FOR PAY?
	Adult female figure (in the household of mother's)—occupation class	OCCUPATION OF ADULT MALE PRESENT IN HOUSEHOLD AT AGE 14 (CENSUS 3 DIGIT)
	Adult male figure (in the household of mother's)—work for pay	DID ADULT MALE PRESENT IN HOUSEHOLD AT AGE 14 WORK FOR PAY?
	Adult male figure (in the household of mother's)—occupation class	OCCUPATION OF ADULT MALE PRESENT IN HOUSEHOLD AT AGE 14 (CENSUS 3 DIGIT)
	Mother's family income (at her young adulthood)	(Available 1979-1986) TOTAL NET FAMILY INCOME IN PAST CALENDAR YEAR (TRUNC)
	Mother's family income (at her young adulthood)—low income status	(Only available in 1980, 1981, 1983, 1984, 1985, 1986) IS TOTAL NET FAMILY INCOME ABOVE OR BELOW THIS LEVEL?

2.A 2 Appendix: Tabulation of the CES-D score

Table 2.A.2 Tabulation of the CES-D score

CES-D score	Freq.	Percent	Cum.
0	1,295	11.2	11.2
1	983	8.5	19.71
2	1,352	11.7	31.41
3	1,794	15.52	46.93
4	1,293	11.19	58.12
5	1,089	9.42	67.54
6	920	7.96	75.5
7	710	6.14	81.64
8	537	4.65	86.29
9	417	3.61	89.89
10	296	2.56	92.46
11	227	1.96	94.42
12	174	1.51	95.92
13	140	1.21	97.14
14	86	0.74	97.88
15	84	0.73	98.61
16	54	0.47	99.07
17	39	0.34	99.41
18	24	0.21	99.62
19	16	0.14	99.76
20	15	0.13	99.89
21	13	0.11	100
Total	11,558	100	

Chapter 3
Physician Response to Pay-for-performance – Evidence from a Natural Experiment

3.1 Introduction

Explicit financial incentives, especially pay-for-performance (P4P) incentives, have been extensively employed and strongly advocated in recent years by health plans and governments in an attempt to improve the quality of health care services. Pay-for-performance is now a concept that is embraced by a lot of policy makers and is deemed to be a critical component of health care reforms. A typical P4P program offers financial rewards to health care providers for meeting pre-established targets for the provision of specific health care services. These explicit financial incentives, which are used within different compensation schemes, aim to motivate health care providers to provide high-quality care.

A variety of P4P programs have been established in several countries. In the United States, as of 2005 at least 100 nationwide P4P initiatives had been sponsored by health plans, employer coalitions and the Centers for Medicare and Medicaid Services (CMS) (Baker and Carter 2005). Initially, most of the P4P programs were targeted at primary care physicians affiliated with Health Maintenance Organizations (HMO). Since 2004 there has been significant

expansion of P4P programs to specialists and hospitals, which use more sophisticated measures for performance assessment (Rosenthal and Dudley 2007, Baker 2004, Baker and Carter 2005). In the United Kingdom, the British National Health Service (NHS) introduced a pay-for-performance contract for family practitioners in 2004 which linked physician income to performance with respect to 146 quality indicators relating to clinical care for 10 chronic diseases, the organization of care and patient experience (Doran et al. 2006). P4P incentive programs have also been used in Canada, Australia, Haiti and other nations (Frolich et al. 2007).

The rationale for employing P4P incentives to induce desired physician behaviour comes primarily from principal-agent theory and incentive-contract theory. The classic principal-agent and incentive contract theories analyze how pay-for-performance can be used to elicit desired behaviours from individuals in the presence of information asymmetry. The analysis focuses particularly on how the ability to elicit desired behaviour is constrained by the noisiness of the performance measures, the extent to which the performance is easily monitored, the ability of agents to handle risk, and the extent to which the desired behaviour consists of multiple tasks (Prendergast 1999; Baker 1992; Hart and Holmstrom 1987; Milgrom and Roberts 1992; Stiglitz 1974). The take-away message from these theories is that performance-based contracting can induce agents to improve performance when payment is based on achieving pre-specified performance targets.

In reality though, using P4P programs to motivate health care providers' behaviour is controversial. Advocates believe that P4P can fix many of the long-standing deficiencies in the health care system, especially the failure to deliver appropriate and evidence-based care to all patient populations. Years of reforms to general payment mechanisms have had little impact on reducing the deficiencies in health care delivery. This has led to the gradual employment of explicit P4P incentives to link financial gains and losses to quality indicators (Maynard 2008). The belief is that, by making payments at least partly contingent on indicators of high-quality care, P4P programs will induce providers to improve health care quality (Rosenthal and Frank 2006). However, critics argue that P4P programs are not as effective as commonly claimed and often create unintended consequences. Some argue that P4P programs can be very costly because payment used to induce even marginal improvements in quality is often expensive (Christianson et al. 2008; Lewis 2009). Others argue that P4P will induce gaming behaviour by physicians such as strategic coding of patient diagnoses, patient selection and patients-exception reporting (Hutchison 2008; Shen 2003; Richards 2009; Doran et al. 2008; Gravelle et al. 2010). Finally, some P4P programs create unintended consequences such as provider focus on the clinical outcomes subject to incentives to the neglect of other aspects of care (Rosenthal and Frank 2006; Mullen et al. 2010).

Theoretical predictions on physician responses to P4P incentives are ambiguous. Health economists generally model physicians as utility-maximizing

service providers who choose their optimal level and mix of services to trade off among income, leisure and other consumption goods (McGuire and Pauly 1991; McGuire 2000). Physician responses to the price increase of the targeted services, generated by individual P4P incentives, are ambiguous because income and substitution effects work in opposite directions. Furthermore, there is no consensus about the specific form of physicians' utility functions. Besides financial objectives, non-pecuniary factors including medical ethics, professional autonomy and social status, and altruistic concerns about patient outcomes are also argued to influence physician utilities (Scott 2001; Eisenberg 1985; Eisenberg 1986). As a result, physicians are less likely to respond to financial incentives when a falling marginal utility of income renders income less attractive in relation to other objectives (McGuire 2000). Moreover, P4P incentives in health care are often embedded within complex compensation systems and provider organizations (Conrad and Christianson 2004; Frolich et al. 2007), where physicians face different incentives from multiple payers and operate in highly regulated settings. The effect of P4P incentives can thus be mitigated by other simultaneous incentives. Therefore, how physicians would respond to P4P incentives remains an empirical issue.

Empirical studies providing convincing evidence of how performance incentives influence physician delivery of targeted services are scarce. Studies based on Randomized Controlled Trials (RCTs) have limited generalizability due to the small scale of the experiments. Although the number of observational

studies is growing, these empirical studies often suffer from poor study-design. Furthermore, the findings from the existing empirical studies are mixed and inconclusive. Most of them find partial effects of P4P incentives in the sense that physicians respond to some of the incentives but not the others; for the subset of incentives which did improve performance, the magnitude of the improvement is modest. A few studies find consistent positive effects but others find no effect. We will discuss these studies, and others, in more detail in the following section.

This study exploits a natural experiment in the province of Ontario, Canada to identify empirically the impact of pay-for-performance (P4P) incentives on the provision of targeted primary care services. The P4P scheme rewards family physicians (FPs) and general practitioners (GPs) when they achieve targeted levels of service provision[29]. Primary care reform in Ontario provides a good setting that allows us to employ a difference-in-differences approach to control for potential sources of bias when identifying the effect of P4P incentives on physician behaviour. The policy intervention exposed some, but not all, of the GPs in Ontario to P4P incentives. Therefore, the GPs who were not eligible for the P4P incentives constitute a natural comparison group for our study design. Also, the timing of the P4P implementation allows us to mitigate perfect confounding of other attributes of primary care reform interventions with P4P. The majority of the GPs were exposed to P4P incentives sometime after they participated in the primary care reforms. Using this group of GPs as the treatment group in a

[29] For ease of exposition, for the rest of the paper I will refer to both FPs and GPs as "GPs".

difference-in-differences with individual fixed effects method allows us to disentangle the impact of P4P incentives from the effect induced by other policy changes.

We exploit an administrative data source which covers the full population of the province of Ontario and nearly all GPs. The administrative databases include detailed information on services provided that constitutes over 98% of all physician activity. By linking different sources of administrative databases, we can observe the group of physicians who were affected by the incentives and the group of physicians that were not affected by the incentives in both pre- and post-intervention periods. The population-based nature of this data provides us with a large sample size, while the rich content of the data allows us to address a variety of potential biases that are caused by "selection on observables" and to partially control for potential biases that are caused by "selection on unobservables".

Furthermore, the universal public insurance and single-payer system in Canada provides an extra advantage for identifying the P4P incentive effects. In multiple-payer settings, such as the U.S., as Robinson notes (Robinson 2001), comprehension and compliance to any payment mechanism will be undermined when physicians face different incentives from multiple insurers or organizations. Therefore, the estimates of the P4P incentives from the US studies are expected to be biased towards zero. In Ontario, however, physicians face only a single payer.

This study also examines the heterogeneity of the P4P incentive effects across different physician types and different practice characteristics. We expect that the

impact of P4P incentives is heterogeneous because both the benefit of responding to P4P incentives and the cost of responding likely differ across physicians, services and practices. We compare the incentive effects across physician age, across practices that differ in patient population size, and across practices with different baseline levels of service provision.

3.2 Empirical Evidence on Physician Response to P4P

A large body of empirical studies has examined the effect of financial incentives on physician behaviour. There is considerable evidence that physicians respond to the incentives embedded in different payment schemes (McGuire and Pauly 1991; McGuire 2000; Hurley et al. 1990; Yip 1998; Nguyen and Derrick 1997; Hickson et al. 1987; Krasnik et al. 1990; Scott and Shiell 1997). There is less evidence on physician responses to explicit financial incentives in the form of targeted performance payments intended to guide specific behaviours.

This study focuses on the effect of pay-for-performance incentives on the behaviour of physicians so we focus the review on thirty studies identified by several recent survey papers (Rosenthal and Frank 2006; Christianson et al. 2008; Petersen et al. 2006; Town et al. 2005; Armour et al. 2001) and by our own search of the literature for papers that pertain to physician responses to P4P incentives (See Appendix 1 for the identified empirical studies). Among the thirty studies,

eight of them are based on RCTs and twenty-two are based on observational studies.

The RCTs examine the effects of alternative forms of performance incentives on the provision of targeted services by physicians, such as bonus, bonus based on capitation payment, and bonus with performance feedback. In most RCTs, the incentives are mostly targeted on preventive care services, including influenza immunizations, mammograms, Pap smear, colorectal screening and pediatric immunization. The sample sizes are generally small.

The results from the RCTs are mixed. Three studies (Grady et al. 1997; Hillman et al. 1998; Hillman et al. 1999) did not detect any significant effect of P4P bonus rewards or bonus rewards combined with performance feedback on physician compliance with cancer screening, pediatric immunization and mammography referrals. Two studies (Fairbrother et al. 1999; Fairbrother et al. 2001) found that a bonus or bonus with performance feedback incentives increased documented coverage levels for childhood immunization, but the measured increase was primarily due to better documentation not better immunization practices. A study of bonus payments for smoking cessation clinics (Roski et al. 2003) found a significant improvement in documentation of patient smoking status and in providing advice to quit, but no effect on quitting rates. The other two RCTs showed a significantly positive effect of using bonus payments at the practice or the clinic level: Kouides et al. (1998) showed that a bonus payment for influenza immunization increased rates by 7 percent; Lawrence et al. (2008)

found that the clinics with P4P payments had higher levels of referral rates on tobacco quitline services than the clinics without payments.

RCTs are often deemed to be the "gold standard" for identifying the causal effects, but these RCTs often suffer from small sample size problems and cannot easily be generalized or extrapolated. All of these RCTs are based on small scale experiments involving fewer than a hundred physicians or practices. One study (Hillman et al. 1998) involved only 52 physician practices in total. As a result, the effect size might not be statistically identified due to lack of power. Moreover, the intervention studied by these RCTs can make it impossible to disentangle the pure P4P financial incentives effects from other quality management tools. Among these RCTs, two studies (Hillman et al. 1998; Hillman et al. 1999) bundled the bonus payment with performance feedback regarding compliance levels; one study (Grady et al. 1997) bundled financial rewards with the provision of education in the form of chart reminder stickers.

The observational studies are mostly based on small to large scale pilot pay-for-quality programs or quality-improvement initiatives adopted by health plans in the U.S., U.K. and Taiwan. These programs generally covered a broader set of quality indicators than merely preventive care services, such as process and outcome measures for diabetic care, asthma and coronary heart disease and other chronic conditions

Doran et al. (2006) evaluated the effect of the nationwide P4P program introduced by Britain's National Health Service in 2004 for family practitioners.

The program linked increases in income to performance with respect to 146 quality indicators covering clinical care for 10 chronic diseases, organization of care, and patient experience. The English family practices attained high levels of achievement meeting the quality indicators, as the median reported achievement was 83.4 percent in the first year of the P4P program (April 2004 through March 2005). But this study is based on cross-sectional analysis so it only established an association between high levels of reported achievement and the P4P contracting, not the real effect of the P4P incentives. As Campbell et al. (2007) noted, because a wide range of initiatives, including limited use of incentive programs, had been introduced in the U.K. since 1990, the high levels of quality attained after the 2004 contract might just reflect improvements that were already under way.

Campbell et al. (2007; 2009) used a before-after design to examine the effect of the 2004 P4P contracting on the quality of care. Both studies measured quality indicators for three chronic conditions --- asthma, coronary heart disease, and type-2 diabetes --- for representative groups of general practitioners. Campbell et al. (2007) measured these quality indicators two times before the P4P contracting (1998 and 2003) and one time after the contracting (2005), and compared the quality score predicted by 1998-2003 trend against the observed quality score in 2005. The results indicate that the introduction of pay-for-performance was associated with a modest acceleration in improvement for two of these three conditions, diabetes and asthma. Campbell et al. (2009) assessed the same quality indicators at an additional time point of 2007, and extended the previous study by

using an interrupted time series analysis. The study found that in 2005 the rate of improvement in quality increased for diabetes care and asthma but remained unchanged for coronary heart disease; by 2007, the rate of improvement for all three conditions had slowed down: as compared with the period before the pay-for-performance scheme was introduced, the improvement rate was unchanged for asthma or diabetes and was reduced for heart disease. Since P4P contracting is offered to all general practitioners in the U.K., neither study could include a plausible control group against which to compare changes in service provision following the introduction of the incentives. Other studies based on the same pay-for-performance scheme in the U.K. (Millett et al. 2007; Steel et al. 2007;Vaghela et al. 2009) examined the effect of P4P incentives on other quality indicators such as smoking cessation and hypertension outcomes, and found statistically significant increases in these quality indicators after the introduction of this P4P scheme. They suffer from the same problem of identification thus fail to provide reliable evidence as they employ only simple before-after analysis.

Evidence of P4P incentives from the U.S. is rapidly growing. Most U.S. studies have been based on small-scale pilot P4P programs adopted by health plans in different states. These studies often suffer from poor study design: some of them only employed simple before-after mean comparison or trend comparison (Levin-Scherz et al. 2006; Young et al. 2007; Cutler et al. 2007; Pearson et al. 2008); others do not provide any counterfactual comparison group (Amundson et al. 2003; Mandel and Kotagal 2007; Chung et al. 2010; Boland et al. 2010; Lester

et al. 2010; Coleman et al. 2007). Some of the programs were targeted at health plans or clinics instead of individual physicians, so the lack of individual-level data makes it difficult to draw inference on physician responses to P4P incentives (Felt-Lisk et al. 2007; Gavagan et al. 2010). Furthermore, results are often limited by the small size of these programs. For example, Beaulieu and Horrigan (2005) examined the effect of performance bonuses on the improvement of nine measures for diabetic care by using only 21 physicians as the treatment group. So it is difficult to draw reliable inference from these studies.

The best evidence to date on the effects of P4P programs are from two observational studies in the U.S. drawn from the P4P initiatives introduced by a large network Health Managed Organization (HMO): PacifiCare Health Plan. The first study (Rosenthal et al. 2005) examined the effect of Quality Incentive Programs (QIP) provided by the PacifiCare Health Plan to medical groups in California in 2002 on physician delivery of cervical cancer screening, mammography and haemoglobin A1c tests. It used a difference-in-differences design by comparing provider groups in California which were affected by these incentives with provider groups in the Pacific Northwest which were unaffected by the incentives but also contracted with PacifiCare Health Plan. It found that outcomes improved for cervical cancer screening, but did not improve for mammography and the haemoglobin A1c test. The second study (Mullen et al. 2009) built on the first paper and examined the effect of QIP incentives along with another larger P4P program by the Integrated Healthcare Association (IHA). It

also concluded that the P4P incentive effects are mixed. In line with the previous study, the analysis found evidence of a positive effect only for cervical cancer screening, but not for mammography, the haemoglobin A1c test and asthma medication. Overall, the study concluded that the pay-for-performance scheme resulted in neither a major improvement in quality nor a notable disruption in care (which some hypothesized would be a negative side-effect).

The findings from these empirical studies suggest that the evidence of physician responses to P4P incentives is mixed and inconclusive. Physicians respond to some P4P incentives but not the others. In general, physicians' response to these financial incentives is of modest size with no evidence of ultimate health improvements for the patients.

3.3 Ontario's Natural Experiment

This study draws on primary care reform interventions in Ontario, Canada as a natural experiment of P4P incentive payments to address the following questions: 1) Does P4P stimulate the delivery of targeted health care services by GPs? 2) Are P4P incentive effects heterogeneous across physician and practice characteristics? Primary care reform in Ontario provided a set of performance-based incentives to some of the primary care physicians in Ontario but not to the others. This produces natural treatment and comparison groups by which to identify the effect of P4P incentives on physician behaviour. The ten-year study period (fiscal years

1998/1999-2007/2008) covers years prior to the provision of the performance-based incentives and those after the implementation. At the beginning of the study period in April 1998, all but a few hundred primary care physicians in Ontario were in the traditional fee-for-service practice; at the end of the study period, more than half of these GPs converted to one or more of the primary care reform models that included the P4P incentives.

3.3.1 Background: Primary Care Reform

Over the last two decades, the province of Ontario, Canada has launched a series of primary care reform (PCR) models to improve the quality of primary health care. The PCR models are intended to improve quality by: 1) providing P4P incentives to stimulate the delivery of targeted health care services; 2) converting from traditional fee-for-service payment to a blended payment method; 3) integrating primary care physicians, nurses and other professionals into more collaborative, multidisciplinary teams (Wilson 2006).

The Ontario Ministry of Health and Long-Term Care (MOHLTC) introduced the different PCR models at different points of time for different purposes. This study focuses on four PCR models: the Family Health Network model (FHN), the Family Health Group model (FHG), the Comprehensive Care Model (CCM) and the Family Health Organization (FHO). The earliest model introduced among these four PCR models is the FHN, which existed as early as 2002, requires a

group practice with at least 3 GPs, and is funded through a blended system of capitation for "core" services provided to rostered patients and fee-for-service for both non-rostered patients and for "non-core" services excluded from the basket of capitated services. FHGs were introduced in 2003, also required a group of 3 or more GPs but the basic payment scheme is an enhanced fee-for-service formula, which consists of the traditional fee-for-service payment for usual care, plus some capitation payments for comprehensive care services provided to rostered patients. The CCM model was introduced in 2005, can include only a solo GP, and is funded through fee-for-service. It is the most similar to traditional fee-for-service (FFS) practice. The FHO model was introduced in 2006, like FHNs, and FHGs requires a group of at least 3 GPs, and is funded through a blend of capitation payment and fee-for-service payment for non-rostered patients and for "non-core" services. FHOs and FHNs are similar in the funding scheme but different in size and rostering regulation. There is no size regulation in patient roster size for the FHO model, but for FHN practices the required minimum roster size is 2,400 patients for a group of 3 GPs while a financial penalty applies if the average roster size is greater than 2,400 patients/GP in the practice. Unlike traditional fee-for-service practice, all of the above four PCR models offer enrolment to their patients (optional for FHGs, required for FHNs, CCMs and FHOs), provide comprehensive care, and impose requirements on GPs to provide a minimum of after-hours care.

3.3.2 Pay-for-Performance Incentives

Ontario initially introduced elements of pay-for-performance in primary care in 1999 to some small-scale pilot PCR models, and expanded it within primary care in 2004. The 2004 Physician Services Agreement included a large number of incentives targeting various aspects of the organization of PCR practices and the care delivered by physicians in those practices. Further, as discussed below, the specific incentives and dates of eligibility differ across the various PCR models.

We focus on a set of P4P financial incentives for five preventive care services (referred to as the Service Enhancement Payments for Preventive Care): Pap smears, mammograms, flu shot for seniors, toddler immunizations, and colorectal cancer screening; and on special payments for services in six areas of care of particular interest to the MOHLTC: payments for obstetrical deliveries, hospital services, palliative care, office procedures, prenatal care, and home visits. Table 1 lists the details of the five performance-based incentives for preventive care services and the six special payments for designated sets of services.

3.3.2.1 P4P Incentives for Preventive Care

The P4P incentives for the five preventive care services include two components: a contact payment and the cumulative preventive care bonus payment. The contact payment rewards PCR practices for contacting patients to schedule an appointment to receive a targeted preventive service. Specifically, the PCR practice receives a contact payment of 6.86 dollars for each eligible patient in the

target population that it contacts and for which it provides the Ministry the required documentation. The cumulative preventive care bonus payment rewards PCR practices for achieving high rates of coverage for the targeted preventive services in the physician's practice populations.

Physicians receive cumulative bonus payments for each service on March 31 of each year based on the proportion of its physicians' eligible and rostered patients who received the targeted service over a specified period of time prior to March 31. Physicians receive a specified amount of money if the proportion reaches a pre-specified coverage threshold, and the payment grows as the proportion exceeds higher thresholds. For example, if 60% of a physician's rostered female patients in the age of 35 to 69 received a Pap smear for cervical cancer screening during the previous 30 months as of March 31, a physician is rewarded 220 dollars. If 65% of the eligible patient population received a pap smear, a physician receives 440 dollars. The physician is compensated with 660 dollars, 1,320 dollars and 2,200 dollars for coverage rates of 70%, 75% and 80%, respectively. It should be noted that, the bonus payment is only based on the proportion of a physician's rostered and eligible patients who received the service in the defined time period; the physician with whom the patient is rostered on March 31 need not have provided this service. For example, if a physician provided a Pap smear to a patient on February 1 and that patient changed physicians on March 1 and rostered with a new physician, the patient's receipt of the Pap smear would count toward the second physician's bonus calculation on

March 31.

It should also be noted that, although the payment is based on the performance of individual physicians, whether the payment is made directly to the individual physician varies across the four PCR models. The payment is made to the physician's PCR practice for GPs in a FHN; how the practice uses the funds received is determined by the practice[30]. Physicians in FHGs, CCMs and FHOs receive the payment directly; it does not go to the PCR practice.

3.3.2.2 P4P Special Payments

The special payments are structured differently. In each case, a physician received a fixed payment if the targeted service was delivered to a minimum absolute level of service provision during the preceding fiscal year, where that minimum is defined in terms of number of services, dollar value of services, number of patients, or a combination of these factors. For each incentive there is also only a single threshold level: if it is reached, the physician receives the special payment; if it is not reached, the physician does not receive the payment. For example, if five or more obstetrical services[31] were delivered to five or more patients in a fiscal year, a physician receives a fixed payment of 3,200 dollars

[30] Beginning in 2006, if there is unanimous agreement among the physicians in a FHN practice, the practice could request that the payments be made directly to its individual physicians rather than the FHN. We have no information on the number of practices that have exercised this option. For ease of exposition, for all incentives we refer to "whether a physician receives a payment" even in those instances when the payment was made to the practice rather than the physician.

[31] Specific services eligible to count toward this special payment include: vaginal delivery, attendance at labor and delivery, Caesarean section, attendance at labor when patient transferred to another centre for delivery, etc.

(with an increase to 5,000 dollars since October 2007). Unlike the preventive care bonuses, the services had to be provided by the physician. Moreover, for all six designated services, the payments were made directly to the physician.

3.3.3 Eligible Physicians

Not all GPs in Ontario were eligible for these financial incentives. In general, these financial incentives were offered only to physicians practicing in a PCR practice. Therefore, physicians who remained in fee-for-service practices were never eligible to receive these P4P incentives. Only physicians who converted from traditional fee-for-service to PCR models were eligible for some or all of the P4P incentives. Furthermore, eligibility of these P4P incentives differs by PCR models. As a result, physicians were eligible for a P4P incentive only after they converted to one of the PCR models and only after the P4P incentives were in effect for the specific PCR model they joined. During the study period of 1999-2008, the P4P incentives were provided at different time points to the four PCR models. Table 2 presents the eligibility timing for the 11 targeted services by PCR model types.

As only some physicians in Ontario were entitled to these P4P incentives, this policy intervention serves as a natural experiment that we can exploit to identify the causal effect of P4P incentives. Since we can observe the practice activities of almost every GP in Ontario over 10 years (1999-2008) and because

this period spans the introduction of P4P incentives implementation, we can assess the impact of P4P incentives within a difference-in-differences framework by comparing the responses of the GPs exposed to the P4P incentives against those not exposed to the P4P incentives.

Of course, the natural experiment formed by this intervention poses some difficulties for identification. First, physicians are not randomly assigned to the PCR models. This will lead to selection bias if we use simple difference-in-differences mean comparison on the responses from eligible GPs against ineligible GPs. Moreover, the PCR model practices are different from the traditional fee-for-service practice in various aspects. Table 3 lists the main differences among each of the four PCR models in the aspects of general payment scheme, practice composition, after-hour services and patient enrolment requirement. Traditional FFS GPs receive only FFS payments, while all PCR model GPs receive a blend of capitation payment and FFS payments, with different proportions of these two components. Unlike the traditional FFS practices, most of the PCR models require GPs to work in group practices (the only exception is CCMs that allow solo practices). Also PCR model GPs have to provide extended services, nurse-staffed telephone health advisory services and on-call services. Lastly, patient enrolment is required in these PCR models except for FHGs but not for FFS GPs. As a result, the identification of the P4P incentive effects may be confounded by differences between the traditional fee-for-service practices and the PCR model practices.

In spite of these problems, the implementation of the performance-based incentives in Ontario still allows us to identify empirically the P4P incentive effects using several identification strategies to mitigate selection bias and control for confounding effects. As described in the methods section below, eligibility for the incentive payments is not perfectly confounded with joining a PCR: some GPs joined a PCR model before they became eligible for bonus payments (unaware that they would later become eligible for such payments). This enables us to distinguish the effects of the incentive payments from the effects of joining a new practice model. Furthermore, variation in general payment schemes and practice settings among the four PCR models themselves provides us an opportunity to disentangle the effect of P4P incentives from that of other primary care reform features.

3.4 Data

3.4.1 Data Sources

The study draws primarily on four administrative databases of the Ontario Ministry of Health and Long-Term Care (MOHLTC), linked by patient encrypted health number and physician encrypted number. The *OHIP Claims Database* provided information on all OHIP-funded services received by each resident of Ontario for each month of the study period; the *Registered Persons Database* provided basic information on each OHIP beneficiary; the *Corporate Provider*

Database provided basic information on each physician and his or her practice; the *Client Agency Program Enrolment* (CAPE) file provided information on the patient roster for each physician in a PCR practice. OHIP claims data allowed us to identify all services provided by every primary care physician in Ontario. The *Client Agency Program Enrolment* (CAPE) data allowed us to match every patient to a physician enrolled in a PCR practice, and to identify if this beneficiary should be counted towards the targeted population for each incentive payment. These data merged with the *Registered Persons Database* provided us with the characteristics of the patient population for each practice. The *OHIP Claims Database* allowed us to construct the yearly utilization rate of each of the targeted services for every physician. The *Corporate Provider Database* allowed us to identify if a GP was enrolled with any of the PCR models at any point in time during the study period. Together these four databases enabled us to construct for each primary care physician in the province of Ontario, a measure of their practice population each year and a record of all services received by those patients during the period of 1998/99 to 2007/08 fiscal years. (See Appendix 2 for all the data sources that we used and the information that we extracted from each source).

3.4.2 Study Sample

The unit of this analysis is a physician. The analyses focus on community-based GPs that do not specialize in a subset of services. We used to following criteria to

select the study sample: (1) include physicians who are GPs throughout the study period; (2) exclude part-time GPs who billed less than 30,000 dollars each year; (3) limit the study sample to GPs in an established practice, we only included physicians who had at least two consecutive years of practice before the study period; (4) include GPs for whom office-based consultations accounted for the majority of their activities; (5) exclude locums as they are not eligible for the bonuses; (6) exclude GPs affiliated with the PCR models for which we do not have sufficient data for the analyses; (7) exclude GPs who converted to FFS for more than one time during the study period for simplicity of the analyses. Table 4 documents how many physicians were excluded by the various criteria when they were applied in the order listed. After applying these criteria, we obtained a core sample of 2,185 GPs.

Since the eligibility scope and implementation dates for the 11 P4P incentives are different for the four PCR models, the composition and the final sample size of the treatment and control groups vary by the P4P incentives. Again for simplicity of the analysis, we dropped physicians whose "treatment" status "turned on" and "off" more than once during the study period[32]. The compositions of control and treatment groups as well as the final sample sizes are presented in Table 5, divided into three subsets of targeted services.

[32] This might be switching back and forth between FFS practice and a PCR model, or switching back and forth between a PCR model which was eligible for the incentives and another PCR model which was not yet eligible for the incentives.

3.4.3 Variable Specification

3.4.3.1 Physician Responses

Physician responses were measured differently for the preventive care services and the designated services for special payments. Because each of the preventive care bonuses is defined with respect to the proportion of a GP's practice population that has received a specified service as of March 31 each year, the outcome variable is defined as the rate of coverage for the relevant period each year for each preventive care service. For the special payments, the outcome variable is defined as the number of services provided or the number of individuals to whom the designated services had been provided.

Analyzing the impact of the incentive payments requires that we identify each GP's practice population on March 31 of each year. Therefore, we used the following steps to define the practice patient population for each GP. For each year we assigned all patients in the Ontario Health Insurance Program (OHIP) physician claims database to a GP and thereby defined a practice population for each GP on March 31 of each year of the study period. Different methods were used to define practice populations for physicians in FFS and physicians in a PCR. Physicians in traditional FFS practice do not roster patients. We defined the practice population for these physicians using the validated methodology developed in Hutchison et al. (1997). Specifically, a physician's practice population is defined as: all individuals for whom the physician billed OHIP for at

least one visit during the previous fiscal year; and all additional patients for whom the physician billed OHIP for at least one visit in each of the two preceding fiscal years. Patients who met these criteria for more than one physician were assigned to the physician who billed for the largest number of visits; if the number of visits was equal, assignment was based on the physician with the most recent visit (for details see Appendix 3). Physicians participating in PCR models have both rostered (the sizable majority) and non-rostered patients. In this case we define the practice population as the set of rostered patients (as indicated by the Ministry Client Agency Program Enrollment database) plus non-rostered patients as assigned by the Hutchison et al. algorithm. As a result, the majority of the OHIP beneficiaries were assigned to a physician for each year of the study period.

After assigning the patients to each physician based on OHIP claims, we counted the number of patients in each physician's practice who received a targeted service during the relevant period and constructed the dependent variables for the empirical analysis for each targeted service for each GP in each year. It should be noted that for the mammogram and senior flu shot bonuses, this study requires additional data because patients can receive these services at specialized clinics whose activity is not captured by the OHIP claims database. For the mammogram bonus, we were able to merge individual-level data on services used in these clinics[33] and so capture all mammograms in the province.

[33] There is a provincial program —Ontario Breast Cancer Screening Program—from which patients can also receive mammograms but these activities were not included in the OHIP claims.

For the senior flu shot bonus, we were not able to do this, so our data exclude such service provisions. This has limited our ability to obtain an unbiased estimate of the P4P incentive effects for this service. We will return to and discuss this limitation in the methods section below.

For the five preventive care bonuses, the dependent variable of each targeted service is defined, as of March 31 each year, as the proportion of a GP's practice population that received the service in question during the relevant period prior to that March 31st. For PCR GPs, this variable is constructed using data from rostered patients only because the Ministry's criterion for payment of the bonus is defined in reference to rostered patients only. We conduct a sensitivity analysis (see Section 6.1.3 below) using an alternate dependent variable that includes both the rostered and non-rostered patients for GPs in PCR models so to obtain a measure that is more consistent across traditional FFS and PCR physicians. A further complication with this dependent variable definition is that PCR physicians can bill a "tracking code" for patients who receive a flu shot at specialized clinics rather than the GP's office, an option not available to FFS physicians. We conduct sensitivity analyses regarding the use of such codes to define flu shot uptake among PCR practices to test the robustness of the findings to this potential problem.

For the six special payments the dependent variable of each designated

---

Therefore, we integrated this part of data provided by Cancer Care Ontario (CCO) into our analysis for mammograms. Unfortunately, we couldn't get any data on flu shots provided in the community flu shot clinics so our flu shot analysis suffers from this data limitation.

service is defined dichotomously, taking on the value of 1 if the physician's service provision met the criteria for the special payment of interest, and 0 if it did not.

3.4.3.2 Independent Variables

As noted above, the *Corporate Provider Database* allowed us to identify if and when a GP joined a PCR practice during the study period. Based on this, we constructed a treated/control dummy indicating if a GP was ever eligible for the incentives during the study period, a pre- and post- dummy indicating if an observation was from a period before or after the implementation of P4P incentives, and a treatment dummy which is an interaction term between the above two dummies, taking on the value of 1 when a GP was eligible for the incentive during the time period in question.

In addition, we included in the analyses a set of independent variables that represent both the supply-side and demand-side characteristics of service utilization. These include characteristics of a physician and the physician's practice, and basic information of the physician's patient population. Physician-specific characteristics are physician age, sex, years in practice, activity level measured by total value of claims submitted each year, and a set of work-load variables including days of work, number of patient visits and number of patient visits per working day. Practice-specific variables include: practice model (FFS, FHN, FHG, CCM and FHO), size of practice population, and a set of practice location characteristics measured by Metropolitan Influence Zone (MIZ)

categories and the Rurality Index of Ontario (RIO). The MIZ categories indicate the degree of influence that metropolitan areas have on the geographic location of a practice; the RIO score indicates the degree of rurality of a practice location. We also control for a set of patient population characteristics, including the mean age of a physician's patient population, and the proportion of female, infant and elderly patients in the practice. The detailed covariate definitions are listed in Appendix 4.

3.4.3.3 Descriptive Statistics of Independent Variables

Table 6 presents sample descriptive statistics at the pre-intervention baseline, defined as of March 31, 2003, disaggregated by the control group and the incentive group.[34] The control group GPs differ at baseline from the incentive group GPs. First, incentive group GPs are younger and have fewer years of practice experience than control group GPs. This observed difference is not surprising because we expect that GPs whose complying costs are relatively smaller are more likely to participate in the PCR models. Younger GPs are more flexible in practice style thus more easily adapt to the specific rules of the PCR practice. Second, a higher proportion of incentive group GPs are female than control group GPs. This might be due to the fact that female GPs are more interested in, or better at, collaborative team production. Third, for all five bonuses and the special payment on palliative care, incentive group GPs worked

[34] As noted above, the definition of the control groups differs slightly across some of the incentives, but the patterns are so similar across the cases that we have collapsed them into one table.

more days and more intensively than the control group GPs before the intervention and they had larger practice sizes. But this pattern is reversed for the other five special payments on obstetrical deliveries, hospital services, office procedures, prenatal care and home visits. For the five bonuses and the special payment on palliative care, the patient population demographics are similar between incentive and control groups, except that the incentive group GPs have practices with slightly more female and infant patients. For the other five special payments, the incentive group GPs also have practices with slightly more female and infant patients, but they also have an older patient population. Finally, incentive group GPs are more homogenous (as indicated by smaller standard deviations) than control group GPs.

3.5 Empirical Methods

3.5.1 Addressing Possible Sources of Bias

As described above, the policy intervention in Ontario serves as a natural experiment that we can exploit to identify the causal effect of P4P incentives. The treatment of interest is a set of P4P incentives targeted on 11 specific health care services or sets of services. Specifically, this policy intervention conditions the eligibility of the P4P incentives on the PCR model-participation status. A simple difference-in-differences approach can provide us with an estimate of the P4P incentive effects by directly comparing the mean change across the PCR model

GPs and the FFS GPs. However, voluntary participation generates non-random assignment of GPs to treatment, invalidating the simple difference-in-differences approach (Meyer 1995). In other words, we expect that the "treated" GPs are systematically different from the "non-treated" ones and these differences may contribute to the observed difference in the response of GPs to P4P incentives. Therefore, identification of the causal effect hinges on how well the selected comparison group represents the counterfactual of the treatment group, and on the extent to which we can mitigate selection bias.

As noted in the descriptive statistics above, GPs who join PCR models differ from those who stay in FFS at the pre-intervention baseline in a number of ways. For example, PCR GPs are younger and have fewer years of practice experience, and their workload is in general different from those in traditional fee-for-service. These differences in physician characteristics might cause estimation bias generated by both "selection on observables" and "selection on unobservables". We discuss in Section 5.2.1 below the empirical strategies used to mitigate selection bias.

One might also be concerned about possible confounding from other factors-- it is possible that some of the observed differences in response to P4P incentives between treatment groups and control groups are actually caused by other unobserved attributes pertaining to the PCR practice rather than the P4P incentives. For example, an important institutional difference between some PCR and FFS practices is that these PCR models are paid by a mixture of FFS and

capitation instead of traditional FFS piece rates. One might expect that FFS physicians respond less to P4P bonuses related to preventive care services because the opportunity cost may be greater for FFS physicians than for physicians paid by capitation or salary in the sense that doing more preventive care may preclude the provision of other services that generate higher fees per unit time. Another type of confounding may arise if we are concerned about separate initiatives that influence the level of utilization of the services being analyzed. The potential sources of this type of confounding and the strategies we use to control for them are described in Section 5.2.2 below.

3.5.2 Identification Strategies

3.5.2.1 Strategies to Mitigate Selection Bias

We employ several identification strategies to mitigate the selection bias that may be generated by both observable and unobservable physician characteristics. First, we control for important aspects of physician characteristics and practice characteristics that might be correlated with the self-selection process and are also important in determining the provision of the targeted services. The data allow us to control for physician characteristics including physician demographics, work experience, and work load measures; and practice characteristics including practice size, geographical location of the practice and patient population characteristics of the practice.

Second, to address selection bias generated by unobservable characteristics, we exploit the longitudinal nature of the data and employ a difference-in-differences approach with individual fixed effects. As noted above, GPs may self-select into PCR models through a process linked to unobserved physician characteristics. This type of selection bias can be reduced to the extent that the unobserved components that determine both the self-selection behaviour and the outcomes are physician-specific and time-invariant, and thereby can be differenced out by a difference-in-differences approach with individual fixed effects.

A potential limitation of the above approach is the lack of control for unobserved temporal individual-specific components that affect the selection into the treatment and control groups (Blundell and Costa Dias 2000). This could be a problem if some GPs self-selected into PCR models because of temporary shocks that are directly related to the targeted health care services. However, this should not be an overriding concern in this study for the following reasons. First, participating in a PCR model is unlikely to depend on short-term changes that affect the utilization rates of the targeted services, such as a sudden demand-side change or an onset of other simultaneous policies that are targeted to these specific services. The monetary values of these P4P incentives constitute a very small proportion of the total income of GPs. So it is unlikely that any temporary changes related to the targeted services caused the conversion behaviour. This assumption is reinforced by the fact that only a very small proportion of GPs who

converted from a FFS practice to PCR models switched back to FFS practice during the ten-year study period. Second, any unobserved temporary shocks that are correlated with PCR participation should not play a major role in determining the utilization of the specific services that are targeted by P4P incentives, because most of the treatment group GPs already converted a number of years prior to becoming eligible for the P4P incentives. Hence, the incentives are unlikely to be the underlying reason for conversion behaviour.

3.5.2.2 Strategies to Control for Confounding Effects

We are concerned about potential confounding from a PCR-practice effect because PCR practices have features (beyond the P4P incentives) not found in traditional fee-for-service practices. We argue that this type of confounding can be controlled in the analyses in the following ways. First, the eligibility timing of the P4P incentives in the PCR models facilitates the reduction of this confounding. The policy intervention provided the P4P incentives to different PCR models in different time periods, but it created essentially three types of physicians groups: a non-incentive group, an incentive group 1 and an incentive group 2 (see Figure 1). The non-incentive group consists of the GPs who remain in FFS over the study period. Since they were never eligible for the incentives, they are used as the legitimate control group in the difference-in-differences design. Incentive group 1 consists of the GPs who joined a PCR model and simultaneously became eligible for the P4P incentives. This group of physicians can be used as part of the treatment group but this is problematic-- given that participation in PCR models is

a voluntary process, the P4P incentive effect is perfectly confounded by the selection into the PCR model for this group of physicians. Incentive group 2 consists of GPs who joined a PCR model before the P4P incentives were introduced and who therefore became eligible for the P4P incentives only after they had participated for some time in a PCR model. This group of GPs pertains to the majority of physicians who were entitled with the incentives in this study. Using this group of physicians as the treatment group can mitigate the problem of confounding. Because these physicians chose to participate in PCR before (and with no expectation for future P4P incentives) the introduction of the P4P incentives, the incentive effect is not perfectly confounded by the other PCR-model features. Second, we use alternative treatment groups in the comparison to mitigate confounding from some specific PCR attributes. This approach is possible for this study since we can exploit the variation on several dimensions across different PCR models to conduct falsification tests on the effects of some specific confounders over the P4P incentive effects. For example, to rule out the possibility that the difference in the general payment scheme is causing the difference in responses, we restrict the treatment GPs to those PCR GPs who were also compensated mainly by fee-for-service and compare their behaviour with the FFS control group GPs. If we still observe a difference in responses, we can conclude that it is not likely that the general payment scheme is causing the observed P4P incentive effects.

Our identification is complicated by potential confounding effects of

separate initiatives that could influence the level of utilization of preventive care services during the study period. Potential confounding from such other initiatives is of greatest concern for senior flu shots, breast cancer screening and colorectal screening. The province has invested heavily in its universal flu vaccination program since 2000, both in making the flu shot available through special clinics and in promoting the uptake of the flu shot. Flu shots obtained through a flu shot clinic rather than in the GP office are not recorded in the OHIP database. Similarly, women can obtain a mammogram through the Ontario Breast Screening Program, which offers specialized clinics for mammograms. Mammograms obtained through these clinics are also not recorded in the OHIP claims database, though, as noted above, we are able to capture such utilization by integrating data from Cancer Care Ontario, the provincial agency that oversees the breast screening program. Finally, beginning in 2004 Ontario launched a pilot program to encourage colorectal cancer screening, and in 2007 launched a population-based colorectal cancer screening program ("ColonCancerCheck") in collaboration with Cancer Care Ontario.

However, none of these initiatives are specific to patients in PCR practices: they offer services to all eligible Ontario residents. Consequently, the inclusion of the fee-for-service control group controls for the general impact of these programs on the receipt of the respective services through GP offices as long as they affected provision equally for physicians in the control and treatment groups. A problem arises only if there is an interaction effect between these programs and

treatment/control status. One concern for flu shots and mammograms is that physicians eligible for incentive payments may have differential incentives to encourage their patients to receive the service through the GP office (and captured by the OHIP database) rather than one of the specialized clinics (not captured by OHIP). Because we capture all mammogram utilization (that included in OHIP and that from Cancer Care Ontario) this does not pose a problem for mammograms. But for flu shots we do not capture shots provided in specialized clinics, and in the presence of a differential incentive, this omission would lead to an over-estimate of the effect of the incentive payment.

Finally, the identification of the difference-in-differences with individual-fixed effects approach is based on the assumption of a parallel trend between treatment and control groups. In order to control for the different time trends across treatment and control groups, we use the difference-in-differences adjusting for differential trends approach as suggested by Bell, Blundell and Reenen (1999). This model relaxes the assumption of parallel trends between the control and treatment group GPs when these differential trends have different impacts on the outcome between the P4P system and the non-P4P system.

3.5.3 Empirical Specifications

We employ the following empirical approaches to evaluate the impact of the P4P incentives.

3.5.3.1 Simple Difference-in-Differences with Pooled OLS

The effect of each P4P incentive can be estimated by comparing the treatment and comparison group in the behaviour change before and after the exposure to the incentives. Consider the model:

$$Y_{itj} = \beta_j' X_{it} + \gamma_j T_t + \rho_j D_i + \delta_j T_t * D_i + \theta_{tj} + \mu_{itj} , \quad (1)$$

where Y_{itj} is the utilization score of service *j* for physician *i* in fiscal year *t*; X*it* is a set of covariates; T*t* is a treatment dummy equal to 1 if this is post-period and 0 otherwise; D*i* is a treatment dummy equal to 1 if this physician is in the treatment group and 0 otherwise; T*t**D*i* is the interaction term taking on a value of 1 if GP *i* was exposed to the P4P incentives at time *t.* The estimated coefficient of this term, δ_j indicates the difference-in-differences (DID) P4P incentive effect. θtj is a set of year dummies; μitj is the idiosyncratic term. The above equation is estimated by a pooled linear or nonlinear panel data model.

In order to account for possible serial correlation of the dependent variable over time, we adjust the standard errors by clustering at the individual physician level in the above simple DID estimation and for all the DID models below. This would mitigate the over-rejection problem for DID estimates (see Bertrand, Duflo and Mullainathan 2004) when the inference of the regular t-statistic is based on unadjusted standard errors[35].

[35] We use the "cluster" option in STATA estimation commands to adjust for standard errors for intragroup correlation among observations over time for each physician. As Bertrand et al. noted (Bertrand, Duflo and Mullainathan 2004), this type of adjustment works well when the number of clusters is large (e.g. N is greater than 50). Our sample size (number of physicians) is sufficiently

3.5.3.2 Difference-in-Differences with Individual Fixed Effects

In order to control for fixed unobserved factors that could influence both selection into a PCR model and provision of the targeted services, we include a set of individual-specific fixed effects:

$$Y_{itj} = \beta_j' X_{it} + \gamma_j T_t + \rho_j D_i + \delta_j T_t * D_i + \theta_{tj} + \varphi_{ij} + \mu_{itj} , \quad (2)$$

where φ_{ij} is a set of physician dummies, and μ_{itj} is the idiosyncratic term. The above equation is estimated by a fixed effects linear or nonlinear panel data model.

3.5.3.3 Difference-in-Differences with Differential Trend Model

To relax the parallel trend assumption we use the difference-in-differences with differential trend model suggested by Bell, Blundell and Reenen (1999). This specification assumes that:

$$\begin{cases} e_{it} = \varphi_i + k_p m_t + \mu_{it} & if \quad T_t * D_i = 1 \\ e_{it} = \varphi_i + k_n m_t + \mu_{it} & if \quad T_t * D_i = 0 \end{cases}, \quad (3)$$

where e_{it} captures the unobservables and the noise. m_t is an unobserved trend. If the P4P GPs and the non-P4P GPs have different trends, the impact of these trends is allowed to differ across the two groups, which is captured by k_p and k_n[36]. This paper follows the regression operationalization of Wagstaff and Moreno-Serra (2009). Incorporating the assumption described in (3), we get the following model:

---

big for this adjustment to mitigate this problem.

[36] Note that P is the subscript for the "P4P" group trend; N is the subscript for the "Non-P4P" group trend.

$$Y_{itj} = \beta_j' X_{it} + \gamma_j T_t + \rho_j D_i + \delta_j T_t * D_i + \varphi_{ij} + k_{nj} m_t + (k_{pj} - k_{nj}) m_t (T_t * D_i) + \mu_{itj}\ , \quad (4)$$

which can be estimated by a fixed effects model including year dummies and year dummies interacted with the treatment dummy, i.e.

$$Y_{itj} = \beta_j' X_{it} + \gamma_j T_t + \rho_j D_i + \delta_j T_t * D_i + \varphi_{ij} + \sum_{\tau=2}^{T} \alpha_{\tau j} Year_\tau + \sum_{\tau=2}^{T} \theta_{\tau j} Year_\tau\ (T_t * D_i) + \mu_{itj}\ . \quad (5)$$

In the above model the impact of P4P incentives varies over time, but the average impact of P4P incentives can be estimated as:

$$\text{Mean P4P impact} = \hat{\delta} + \frac{\sum_{\tau=2}^{T} \theta_{\tau j}}{T-1}\ . \quad (6)$$

Because the parallel trend assumption implies $k_p = k_n$, this assumption can be tested by testing the nonlinear restriction:

$$\frac{\sum_t m_t (k_{pj} - k_{nj})}{\sum_t m_t\, k_{nj}} = \frac{(k_{pj} - k_{nj}) \sum_t m_t}{k_{nj} \sum_t m_t} = \frac{\sum_{\tau=2}^{T} \theta_{\tau j}}{\sum_{\tau=2}^{T} \alpha_{\tau j}} = 0\ . \quad (7)$$

Non-rejection of the hypothesis would suggest that $k_{pj} = k_{nj}$ and provide evidence in favour of the parallel trend assumption and the difference-in-differences with individual fixed effects model.

It should be noted that we could only run a full set of regression analyses described above for the five bonuses, but not for the six special payments. As indicated by Table 5, because the eligibility scope and implementation dates for the 11 P4P incentives are different across the four PCR models, the composition

and the final sample size of the treatment and control groups vary by the P4P incentives. Most of our P4P GPs became eligible for the five preventive care bonuses in 2006, except colorectal cancer screening, for which most of our P4P GPs became eligible in 2005. For the six special payments, most of the P4P GPs became eligible in 2005, 2006 and 2007, except for the palliative care special payments, for which most of the P4P GPs became eligible in 2003, 2004 and 2005. Accordingly, for the six special payments, we estimate the difference-in-differences models separately for three subsets of P4P GPs based on the year they became eligible for the payments. Moreover, as the five bonuses were provided to all four PCR models considered in the study while the six special payments were provided to only some of the PCR models (i.e. FHNs and FHOs), there are far fewer GPs constituting the treatment groups in the analysis for the six special payments than for the five cumulative bonuses. As a result, we could estimate the full set of difference-in-differences models and conduct the robustness checks and sensitivity analyses for the five bonuses, but could only estimate the simple pooled difference-in-differences model with the full sample for the six special payments. For the same reason, we could only conduct subgroup analyses for the five bonuses, but not for the six special payments.

3.6 Empirical Results

3.6.1 Descriptive Trends of Physician Responses

We can only document the extent to which GPs contacted patients to arrange the receipt of preventive services for the period after the identifying codes were introduced in the fee schedule. Table 7 presents the proportion of eligible physicians who submitted at least one claim for contacting a patient to arrange an appointment to deliver a preventive care service. Two things are noteworthy: (1) the rate of uptake is relatively low — with the exception of a couple of years for the senior flu shot, less than 45% of eligible physicians submitted even a single claim; (2) and there is no noticeable upward trend — in fact, the proportion has been falling in recent years for 4 of the 5 services. Figure 2 presents the mean number of claims per eligible physician for each service. For all services except senior flu shots, the mean number of claims per eligible physician is fewer than 20; even for senior flu shots the mean exceeds 100 for only one year during this period. Overall, there appears to have been little response to these contact incentive payments.

The main outcome measures for this study are the utilization rates of the services that are targeted by the 11 P4P incentives. The unadjusted time paths of compliance levels of all the targeted services are shown in Figure 3 through Figure 13. The horizontal axis represents the years from March 31, 1999 to March 31, 2008. For the preventive care services, the vertical axis is the mean proportion of patients who received the targeted services. For the special payments the vertical axis is the proportion of physicians who achieved the targeted

performance level. The lines represent the time trends for the control group, and for treatment groups defined in terms of the year when a GP first became eligible for the incentive. For example, incent2003 represents GPs who were first eligible during fiscal year 2002-2003. We can detect a specific pattern of change in trend to the introduction of the P4P incentives for Pap Smears, colorectal cancer screenings, and palliative care. For these services, compared to the control group, provision in the incentive groups started to increase and diverge at the time of exposure to the incentives. This suggests possible effects of the P4P incentive payments for these services. The trend for mammograms displays equivocal evidence. We could not detect any specific pattern in the trend for senior flu shots, toddler immunizations, obstetrical deliveries, hospital services, office procedures, prenatal care and home visits. For most of these 11 services, incentive group GPs started with higher baseline compliance levels so it is possible that a selection effect exists in the means of the compliance levels for treatment versus control groups.

3.6.2 Estimation Results for the Preventive Care Bonuses

3.6.2.1 Estimates for the Full Sample

Table 8 presents the estimates of the P4P incentive effects for the five preventive care bonuses based on three difference-in-differences models for the full sample. Column (a) lists the baseline compliance level of each targeted service, which is

defined as the average utilization rate of this service in 2003. Panels (b), (c) and (d) of Table 8 present the estimates of the P4P incentive effects based on a difference-in-differences with the pooled OLS model, a difference-in-differences with individual fixed effects model, and a difference-in-differences with differential trend model, respectively. The marginal effects estimates indicate the percentage change of the service provision due to the introduction of each bonus payment. In order to account for possible correlation of the observations over time for each physician, we calculated the robust standard errors by clustering by individual physician. The results based on a difference-in-differences with the fixed individual effects model show that the bonus payment had a statistically significant effect on the provision of senior flu shots, Pap Smears, mammograms and colorectal cancer screenings, while its effect on the provision of toddler immunizations is not statistically significant[37]. The absolute level of increase in compliance is 2.8%, 4.1%, 1.8% and 8.5% for senior flu shots, Pap smears, mammograms and colorectal cancer screenings, respectively. It is notable that the marginal effect estimates based on a difference-in-differences with the pooled OLS model are similar to the above figures, except that the incentive effect

[37] As noted by Moulton (1986, 1990), and Donald and Lang (2007), in regression models with mixture of individual and grouped data, the failure to account for the presence of common group errors can generate estimated standard errors that are biased downward dramatically. Our DID estimates may suffer from this problem, under the assumption that physicians in the same practice may have correlated standard errors. Accordingly, adjusting the standard errors for clustering by practice (instead of clustering by individual physician) could correct for the over-rejection problem with our DID estimates. However, we have not done so because we could not properly identify physician practice for FFS GPs and CCM GPs. As a result, our current estimates overstate the statistical significance of the P4P incentive effects, and adjustment for clustering would weaken the evidence of an incentive effect.

estimates are not significantly different from zero for senior flu shots and mammograms in the pooled OLS model. The estimates based on a difference-in-differences with the differential trend model are consistent with those from the individual fixed effects model while indicating slightly larger effects for all five services.

Panel (e) of Table 8 presents the test statistic and the p-value from the nonlinear restriction test on the parallel-trend assumption as described in Equation (7). The results indicate that the null hypothesis of a common trend between the P4P GPs and non-P4P GPs is not rejected at the 5% level for senior flu shots, toddler immunizations and mammograms but is rejected for Pap smears and colorectal cancer screenings. Therefore, the parallel trend assumption is reasonable for senior flu shots, toddler immunizations and mammograms. For these services, the regression results from the difference-in-differences with fixed individual effects model are plausible, while we prefer the results from the differential trend model to the results from the difference-in-differences with fixed individual effects model for Pap smears and colorectal cancer screenings. Even though the parallel trend assumption does not hold for some services, the magnitude of the estimated incentive effects is similar across these two models.

3.6.2.2 Robustness Checks with Alternative Samples

The above results may be subject to bias due to the confounding from other PCR model characteristics. As a robustness check we restrict the treatment group to PCR GPs who joined PCR practices before becoming eligible for P4P incentives.

This robustness check is conducted only for the five preventive care bonuses. Panel (b) in Table 9 presents the estimated P4P incentive effects from a difference-in-differences with individual fixed effects model based on the sample of GPs who joined PCR practices before becoming eligible for the P4P incentives. Column (a) lists the baseline compliance level of each targeted service in 2003 for this sample. The regression results show that the estimated P4P incentive effects are robust to this refinement of the treatment group. Therefore, we conclude that the estimated P4P incentive effects are unlikely to be generated by other PCR practice characteristics.

We also used a falsification test to check whether the observed incentive effects are linked to the general payment scheme, which differs between most treatment and control physicians. As a second robustness check we restrict the treatment group to GPs working in PCR practices that are paid primarily by FFS. If we observe that the response of this subgroup of GPs is not significantly different from those in traditional FFS practices, we have evidence that compromises the estimated P4P incentive effects. Panel (d) of Table 9 presents the estimates of P4P incentive effects based on the sample of GPs in PCR models funded primarily by FFS. The results indicate that refining the treatment group in this way does not change the estimates of the incentive effects from those based on the full sample. This is consistent with the full sample estimation and first robustness check. Overall, the results from these two robustness checks indicate that the full-sample estimates do not suffer from the possible confounding of other

PCR attributes.

3.6.2.3 Sensitivity Analyses on the Study Design

In order to test the sensitivity of the regression results, we conducted four sensitivity analyses to address limitations on the study design and assumptions.

As noted above, one complication with the current study design is that PCR physicians can bill a "tracking code" for patients who receive a flu shot at specialized clinics rather than the GP's office, an option not available to FFS physicians. We conduct sensitivity analyses regarding the use of such codes to define flu shot uptake among PCR practices to test the sensitivity of the findings to this potential problem. In the first sensitivity analysis, we redefined the dependent variables for the incentive payments by excluding the claims from the PCR physicians with these shadow-billed tracking Q-codes. This would give us a lower bound of the true estimates of the incentive effects. Table 10 presents the estimated marginal effects based on a difference-in-differences with individual fixed effects model for the full sample and the two alternative samples we used above. The results from this sensitivity analysis indicate that the base case estimates are robust to this change of definition in that the significance level of the incentive effects stays the same while the magnitude is slightly smaller (as we would expect). The basic conclusions from the main analysis therefore appear to hold.

The second sensitivity analysis aims to test for the consistency of different methods we used to calculate the performance level for the FFS physicians and

the PCR physicians. As discussed previously, the dependent variable could not be defined identically for the PCR and the FFS GPs: for PCR GPs, this variable is constructed using data from rostered patients only because the Ministry's criterion for payment of the bonus is defined in reference to rostered patients only. Therefore, we added in the non-rostered patients for the PCR physicians in the calculation of the dependent variables and compared the performance level based on both rostered and non-rostered patients against that of the FFS physicians. Table 11 presents the estimated marginal effects for this sensitivity analysis based on a difference-in-differences with individual fixed effects model for the full sample and the two alternative samples we used above. The results show that the estimates are robust to this change in the dependent variable definition. The magnitude of the estimated incentive effect is slightly larger under the second sensitivity analysis. We interpret this as the selection effect introduced by our algorithm of assigning the patients—the non-rostered patients within a PCR physician's patient population that are assigned by our algorithm are utilizers of the services thus are more likely to receive preventive care services from this physician.

One additional rule for the FHG and CCM physicians to receive the bonus payment is that they need to reach a minimum threshold of patient roster size. In the third sensitivity analysis, we estimate the incentive effects over only the subset of FHG and CCM physicians with rosters over the minimum threshold at the time of the bonus introduction. The results from this sensitivity analysis represent

short-run responses as opposed to long-run responses represented by the base case results, because the patient roster size could be endogenized over time. Table 12 presents the estimated marginal effects for the third sensitivity analysis based on the difference-in-differences with individual fixed effects model for the full sample and the two alternative samples we used above. The results from this sensitivity analysis confirm the significance of the incentive effects and show a slightly larger response from the physicians who had already achieved the minimum roster size.

At the beginning of the bonus payment introduction, physicians might be just starting to enroll patients into their practice or ramping up for the incentive payment. The calculated performance level might therefore be noisy in the sense that the targeted performance is based on the proportion of patients who received the services. In the fourth analysis we dropped the observation of transition year, i.e. the first year that the treatment group GPs became eligible for the bonuses, from the empirical analyses to remove the potential noise in the observed behaviour in the first period of transition. Table 13 presents the estimated marginal effects for this sensitivity analysis based on the difference-in-differences with individual fixed effects model for the full sample and the two alternative samples we used above. The estimated marginal effects of incentives are not sensitive to this change and are slightly larger in size than the base case results.

3.6.3 Estimation Results for the Special Payments

Table 14 presents the P4P incentive effects for the six special payments from estimating the difference-in-differences with pooled logit model. Column (a) lists the baseline compliance level of each targeted service, which is defined as the proportion of GPs whose pattern of service provision in 2003 exceeded the special payment target level. We present the results separately for the subsets of GPs who became eligible for these incentives in 2005, 2006 and 2007[38] in panel (b), (c) and (d) respectively. The marginal effect estimate indicates the absolute change in the proportion of physicians whose service provision is predicted to exceed the target level as a result of the special payments. There is no statistically significant incentive effect that is consistent over these three subsamples for any of these special payments[39]. Overall, the results suggest little if any response to the special payments: all the estimates are small and not statistically different from zero.

3.6.4 Estimates of the P4P Incentive Effects for Subgroup Analysis

There are reasons to expect that responses may differ by physician age, practice size, and baseline level of compliance. To investigate this we conduct three sets of subgroup analyses and present the results in this section[40]. The subgroup

[38] For the palliative care payment, the results are presented separately for the subsets of GPs who became eligible for the incentive since 2003, 2004 and 2005 instead.

[39] Statistically significant results are found only for one subsample for office procedures and prenatal care.

[40] A hypothesis that response is associated with the size of the group a GP works in cannot be

analyses are conducted only for the five bonuses due to the small sample sizes for the six special payments.

The first panel of Table 15 presents the estimates from the difference-in-differences with individual fixed effects model for the subgroup analyses by physician age for the five preventive care bonuses. We see a clear age gradient for Pap smear, mammograms and colorectal cancer screenings whereby younger physicians respond more to the P4P bonuses than do older physicians. An age-gradient is not discernable for senior flu shots and toddler immunizations. For senior flu shots, we observe that only middle-age physicians responded to the incentives. This indicates the possibility that the relatively weak incentive effect from the whole sample analysis for senior flu shots is driven by the response of the middle-age physicians. We only detect a statistically significant incentive effect in the oldest age group for the toddler immunization bonus, but this effect is only weakly significant at the 10% level.

The second panel of Table 15 presents estimates by practice size. Overall, the results indicate that physicians with larger practices tend to be more responsive to the P4P incentives. For the Pap smear incentive there is essentially no difference in the magnitude of the estimated effects across categories of practice size. For the mammogram and senior flu shot incentives, there is a statistically significant incentive effect only for the largest practices but no effect for small-size or mid-size practices. For colorectal cancer screenings, the pattern

tested with our data due to the missing information on group size for the control group GPs.

is clear that the incentive effect is larger for larger sized practices.

The third panel of Table 15 presents the difference-in-differences estimates for the subgroup analysis by baseline level of compliance. For three of the five preventive services, the response is greatest for those physicians with the lowest levels of baseline provision (senior flu shots and mammograms) or for physicians with the lowest and middle levels (colorectal cancer screenings)[41]. This is consistent with the hypothetical pattern that physicians with lower baseline compliance levels tend to respond more, except for Pap smears in which physicians in the middle quartiles responded the most.

3.7 Discussion and Conclusions

Our estimates of the incentive effects indicate that the cumulative preventive care bonus payments for Pap smears, mammograms, senior flu shots and colorectal cancer screenings have modestly improved the performances of GPs in the provision of these targeted services. The bonus on toddler immunizations and the special payments on obstetrical deliveries, hospital services, palliative care, office procedures, prenatal care and home visits, had no effect on the provision of these targeted services. Regression results are consistent and similar in magnitude across the series of difference-in-differences models that we use in sequence to

[41] Note that for mammograms, the incentive effects for the middle-quartile and top quartile are about the same; while for colorectal cancer screenings, the incentive effects for the lowest quartile and the middle quartile are about the same.

partly control for "selection on observables" and "selection on unobservables". The results from the robustness checks with alternative study samples suggest that it is unlikely that the baseline estimates are driven by confounding effects between P4P incentives and other features of PCR practices. The sensitivity analyses also indicate that the main regression results are robust to different definitions of dependent variables and alternative estimation samples, reassuring the validity of our study design.

In general, our empirical results confirm the empirical literature, which indicates little effect of employing P4P incentives to improve the quality of health care. Among the eleven incentives we considered, seven of them did not result in significant improvement of service provision, while the other four only slightly increased the utilization of the targeted services. As noted above, because we could not adjust the standard errors for possible clustering on practice in estimating the difference-in-differences models, our current estimates overstate the statistical significance of the P4P incentive effects, reinforcing the general conclusion that these incentives are not very effective. Unlike evidence based on P4P programs in the U.S., our findings are derived from observations in a publicly funded single-payer system, so the effect of P4P incentives is not confounded by the institution of multiple payers. Moreover, we found that even for the incentives that generated responses, the magnitude of the response rates varies across targeted services, across physician characteristics and across practice settings. Specifically, physicians responded to the bonuses for preventive care services but

not for special payments. Physician responses differ significantly across physician age and initial service provision level.

The differing physician responses to the preventive service bonuses and the special payments may be due to a number of factors. First, the costs of complying are different for these two sets of incentives. The preventive care services targeted by the five bonuses do not require special costs in provision and are within the expected competency of a GP. Some preventive care services can even be provided by non-physician staff. However, the services targeted by the six special payments often require a fixed cost. Services like obstetrical deliveries often incur some cost related to insurance premiums and require a commitment to be available for deliveries at all hours of the day. Also, providing services like hospital visits and home visits involves additional time costs and re-organization costs (i.e. re-organizing one's schedule for visits). Therefore, relatively larger financial incentives are required to cover these costs and to induce desired behaviours by GPs. Second, providing preventive care is well documented to be effective and is well established as consistent with high-quality care, but services subject to special payments in Ontario have no strong link to quality of care. Lastly, the preventive care bonuses are complementary to other attributes of the PCR models while this is not true for the special payments. For example, unlike the physicians remaining in FFS practice, the physicians who participated in PCR models were eligible for financial support to adopt electronic medical record systems that can provide automatic reminders when a patient should receive

regularly scheduled services. This feature generally does not apply to the provision of special services as it does to the provision of preventive care services.

The general take-away message from our empirical results is that physicians do not automatically respond to performance-based financial incentives as expected. Although principal-agent theory suggests the potential to employ P4P incentives to motivate physicians in providing high-quality care, physician responses to such incentives are not easily predicted. The heterogeneity of physician responses found in our study suggests that physician behaviours may be constrained by a complex set of objectives that we do not directly observe. Therefore, more refined positive analyses on physician health care delivery are warranted to inform future implementation of different and customized incentive schemes to elicit desired physician behaviours. Overall, our results deliver a cautionary message regarding the effectiveness of employing pay-for-performance to increase the quality of health care. The overall small physician responses to the introduction of P4P incentives in Ontario indicate the rather low power of using these incentives to motivate high quality care. One possible reason is that the absolute size of the financial incentives for these services in general is too small to generate the desired response from the physicians. After all, the income increase related to these incentives is only a small proportion relative to the total income for most of the GPs, so the increase of marginal utility related to this income increment likely works very marginally in a physician's service provision decisions. Nonetheless, we would then expect that it will be even more costly to

achieve the pre-specified improvement of service provision if we continue to employ the same incentive structure. As indicated in the recent literature on pay-for-performance, the P4P incentives need to be more carefully designed (Christianson et al. 2008; Epstein 2006; Hutchison 2008). As noted above, the cost of complying may vary substantially among different types of procedures and services. Therefore, tailoring the absolute size of financial incentives for different targeted services according to the relative costs of complying may provide a more cost-effective solution. Furthermore, our findings also suggest that there is only limited scope of using P4P incentives to increase the provision of targeted services. The employment of P4P incentives is only effective when the targeted performance or tasks are strongly linked to professional standards of high quality care. This is reflected in the fact that physicians tend to be more responsive to P4P incentives designed around the provision of preventive care services, which are unquestionably consistent with medical guidelines on providing high-quality care. Therefore, future implementations of P4P incentives could be restricted only to these services. Finally, the P4P incentives should be redesigned so that the target measures are more closely related to real standards of high quality care. For example, financial incentives can be linked to quality indicators that aim to increase access to health care, or to those representative of evidence-based health care.

Further studies on performance payment incentives can be extended in several directions. Like much of the current literature on P4P, we could not obtain

any patient health outcome measures, so we only rely on utilization rates or provision levels in our analyses. These measures may not be representative of the health care quality *per se* and patient health outcomes would be better indicators of quality. Therefore, it will be important to document the effect of P4P incentives on patient health outcomes if such data become available in the future. Moreover, it is interesting to test whether there is a "spill-over" effect of only rewarding the provision of a small subset of services. It is possible that physicians reallocate their time or other resources from the unrewarded services to the rewarded services to obtain more income. Finally, exploring other factors that might be complementary to P4P incentives will help us design better P4P programs for eliciting optimal behaviour among physicians.

References

Amundson, G., L. I. Solberg, M. Reed, E. M. Martini, and R. Carlson. (2003) "Paying for quality improvement: Compliance with tobacco cessation guidelines", *Joint Commission Journal on Quality and Safety* 29 (2): 59-65.

Armour B, Pitts M, Maclean R, et al. (2001) "The Effect of Explicit Financial Incentives on Physician Behaviour," *Archives of Internal Medicine*;161:1261-6.

Baker, G. (1992) "Incentive contracts and performance measurement". *The Journal of Political Economy*, Vol. 100. no. 3, pp. 598-614

Baker G. (2004) "Pay for performance incentive programs in healthcare: market dynamics and business process". San Francisco, CA: Med-Vantage;

Baker G., Carter B. (2005) "Provider Pay-for-performance incentive programs: 2004 National Study Results". San Francisco, Calif: Med-Vantage;

Beaulieu, ND, Horrigan, DR. (2005) "Putting smart money to work for quality improvement." *Health Services Research*. 2005 Oct; 40(5 Pt 1):1318-34

Bell, B., Blundell, R., Van Reenen, J. (1999) "Getting the unemployed back to work: the role of targeted wage subsidies." *International Tax and Public Finance*, 6 (3), 339-360.

Bertrand M., Duflo E. and Mullainathan S., (2004) "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, vol. 119(1), pages 249-275, February.

Blomqvist, A., (1991). "The doctor as double agent: Information asymmetry, health insurance, and medical care," *Journal of Health Economics*, Elsevier, vol. 10(4), pages 411-432

Blundell, R., Costa Dias, M. (2000) "Evaluation Methods for Non-experimental Data." *Fiscal Studies*, 21 (4), 427-468.

Boland GW, Halpern EF, Gazelle GS. (2010) "Radiologist Report Turnaround Time: Impact of Pay-for-Performance Measures" *AJR Am J Roentgenol.* 2010 Sep;195(3):707-11.

Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. (2007) "Quality of primary care in England with the introduction of pay for performance", *New England Journal of Medicine* 2007 Jul 12; 357(2):181-90.

Campbell S, Reeves D, Kontopantelis E, Sibbald B, and Roland D.M. (2009)

"Effects of Pay for Performance on the Quality of Primary Care in England" *New England Journal of Medicine*; 361:368-378

Christianson, J. B., Leatherman S., Sutherland K. (2008). "Lessons From Evaluations of Purchaser Pay-for-Performance Programs: A Review of the Evidence." *Medical Care Research and Review*, 65(6_suppl): 5S-35.

Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS.(2010) "Effect of physician-specific pay-for-performance incentives in a large group practice." *Am J Manag Care.* 2010 Feb 1;16(2):e35-42.

Coleman, K., Reiter, K. L., Fulwiler, D. (2007) "The Impact of Pay-for-Performance on Diabetes Care in a Large Network of Community Health Centers" *Journal of Health Care for the Poor and Underserved*, Volume 18, Number 4, November 2007, pp. 966-983

Conrad, D. A., Christianson, J. B. (2004) "Penetrating the "Black Box": Financial Incentives for Enhancing the Quality of Physician Services." *Medical Care Research and Review*, 61 (3), 37S-68S.

Cutler TW, Palmieri J, Khalsa M, Stebbins M.(2007) "Evaluation of the relationship between a chronic disease care management program and california pay-for-performance diabetes care cholesterol measures in one medical group" *J Manag Care Pharm.* 2007 Sep;13(7):578-88.

Donald S. and Lang K., (2007) "Inference with Difference-in-Differences and Other Panel Data," *Review of Economics and Statistics*, vol. 89(2), pages 221-233, 03.

Doran, T., C. Fullwood, et al. (2006). "Pay-for-performance programs in family practices in the United Kingdom." *The New England Journal of Medicine,* 355(4): 375-84.

Eisenberg JM. (1985) "Physician utilization: The state of research about physicians' practice patterns." *Med Care*. 23:461-83.

Eisenberg JM. (1986) "Doctors' Decisions and the Cost of Medical Care". Ann Arbor, MI: Health Administration Press

Ellis R. and McGuire T. (1990) "Optimal payment systems for health services" *Journal of Health Economics*, vol. 9, issue 4, pages 375-396

Epstein, A. M. (2006) "Paying for performance in the United States and abroad," *New England Journal of Medicine*, 355:406-8.

Fairbrother, G., K. L. Hanson, S. Friedman, and G. C. Butts. (1999), "The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates", *American Journal of Public Health* 89 (2): 171-175.

Fairbrother G, Siegel MJ, Friedman S, Kory PD, Butts GC. (2001) "Impact of financial incentives on documented immunization rates in the inner city: results of a randomized controlled trial", *Ambulatory Pediatrics* 2001 Jul-Aug; 1(4):206-12.

Felt-Lisk S, Gimm G and Peterson S. (2007) "Making Pay-For-Performance Work In Medicaid" *Health Affairs*, 26, no. 4 : w516-w527

Frolich, A., Talavera, J. A., Broadhead, P. and Dudley, R. A. (2007) "A behavioural model of clinician responses to incentives to improve quality." *Health Policy*, 80 (1), 179-193.

Gavagan T, Du H, Saver B, Adams G, Graham D, McCray R, and Goodrick K (2010) "Effect of Financial Incentives on Improvement in Medical Quality Indicators for Primary Care" *J Am Board Fam Med*, September-October 2010; 23: 622 - 631.

Gosden, T., F. Forland, et al. (2000). "Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians (Cochrane Review)." *Cochrane Database of Systematic Reviews*(3).

Gosden, T., F. Forland, et al. (2001). "Impact of payment method on behaviour of primary care physicians: a systematic review." *Journal of Health Services Research & Policy* 6(1): 44-55.

Grady KE, Lemkau JP, Lee NR, Caddell C. (1997), "Enhancing mammography referral in primary care", *Preventive Medicine,* 1997 Nov-Dec; 26(6):791-800.

Guido W. Imbens, (2004) "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, MIT Press, vol. 86(1), pages 4-29, 06.

Heckman, J., Ichimura, H., and Todd, P., (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, Blackwell Publishing, vol. 64(4), pages 605-54, October.

Hickson, G., Altemeier, W., and Perrin J., (1987) "Physician Reimbursement by Salary or Fee-for-Service: Effect on Physician Practice Behaviour in a Randomized Prospective Study," *Pediatrics*, Vol. 80, No. 3, pp. 344-350

Hillman, AL, Ripley, K, Goldfarb, N, Nuamah, I, Weiner, J, Lusk, E. (1998)

"Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care." *American Journal of Public Health*, *88* 1699-1701.

Hillman, A. L., K. Ripley, N. Goldfarb, J. Weiner, I. Nuamah, and E Lusk. (1999), "The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care", *Pediatrics* 104 (4): 931-935.

Hurley, J., R. Labelle and T. Rice (1990), "The relationship between physician fees and the utilization of medical services in Ontario", *Advances in Health Economics and Health Services Research* 11:49-78

Hurley, J., DeCicca,P., Li J. and Buckley G. (2011), "The response of Ontario primary care physicians to pay-for-performance incentives." Final Report (Janurary 26,2011) submitted to the Ontario Ministry of Health and Long-Term Care (MOHLTC)

Hutchison, B. (2008). "Pay for Performance in Primary Care: Proceed with Caution, Pitfalls Ahead (Editorial)." *Healthcare Policy*, 4(1): 5.

Hutchison, B., S. Birch, et al. (1996). "Do physician-payment mechanisms affect hospital utilization? A study of Health Service Organizations in Ontario." *CMAJ*, 154(5): 653-661.

Hutchison, B., J. Hurley, et al. (1997). "Defining the Practice Population in Fee-for-Service Practice." *Health Services Research*, 32 (1): 55-70.

Kouides, RW, Bennett, NM, Lewis, B, Cappuccio, JD, Barker, WH, LaForce, FM. (1998) "Performance-based physician reimbursement and influenza immunization rates in the elderly. The primary-care physicians of Monroe County." *American Journal of Preventive Medicine*, (14), 89-95.

Krasnik, A., P .P. Groenewegen , P.A. Pedersen, P. von Scholten , G. Mooney, A. Gottschau, H.A. Flierman and M.T. Damsgaard (1990), "Changing remuneration system: effects on activity in general practice", *British Medical Journal*, 300 (6741), 1698-701

Lawrence C. A., James H. B., Steven S. F., Nina L. A., Colleen M. K., Bruce A. C., William S. N., Mark E. L., Jasjit S. A., Marc W. M.,(2008) "A Randomized Trial of a Pay-for-Performance Program Targeting Clinician Referral to a State Tobacco Quitline", *Arch Intern Med.* 2008;168(18):1993-1999.

Lester H, Schmittdiel J, Selby J, Fireman B, Campbell S, Lee J, Whippy A, Madvig P. (2010) "The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators" *BMJ*; 340:c1898

Levin-Scherz J, DeVita N and Timbie J. (2006) "Impact of Pay-for-Performance Contracts and Network Registry on Diabetes and Asthma HEDIS Measures in an Integrated Delivery Network" *Medical Care Research and Review,* 2006 February vol. 63 no. 1 suppl 14S-28S

Lewis, S. (2009). "Pay for Performance: The Wrong Time, the Wrong Place?" *Healthcare Quarterly*,12(3):8-9.

Ma, Ching-to A. & McGuire, T. (1997). "Optimal Health Insurance and Provider Payment," *American Economic Review*, vol. 87(4), pages 685-704

Mandel K, Kotagal U. (2007) "Pay for Performance Alone Cannot Drive Quality." *Arch Pediatr Adolesc Med.* ;161(7):650-655.

Maynard, A. (2008). "Payment for Performance (P4P): International experience and a cautionary proposal for Estonia", World Health Organization Europe - Health Financing Policy Paper, Division of Country Health Systems: 30.

McGuire, T. G., and M.V. Pauly (1991), "Physician response to fee changes with multiple payers", Journal of Health Economics 385-410

McGuire, Thomas G., (2000) "Physician agency", Handbook of Health Economics, in: A. J. Culyer & J. P. Newhouse (ed.), Handbook of Health Economics, edition 1, volume 1, chapter 9, pages 461-536 Elsevier.

Meyer, B., (1995) "Natural and quasi-experiments in economics" *Journal of Business and Economic Statistics*, vol. 13, 2 (April) pp. 151-161Millett C, Gray J, Saxena S, Netuveli G and Majeed A, (2007) "Impact of a pay-for-performance incentive on support for smoking cessation and on smoking prevalence among people with diabetes" *CMAJ*, June 5, 2007; 176 (12).page 1705-1710

Moulton, Brent R., (1986)"Random group effects and the precision of regression estimates," *Journal of Econometrics*, vol. 32(3), pages 385-397, August.

Moulton, Brent R, (1990) "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit," *Review of Economics and Statistics*, vol. 72(2), pages 334-38, May.

Mullen K. J., Frank R.G., Rosenthal M. B. (2010) "Can You Get What You Pay For? Pay-For-Performance and the Quality of Healthcare Providers", *RAND Journal of Economics*, 41: 64–91

Newhouse, J. (1996). "Reimbursing Health Plans and Health Providers: Efficiency in Production versus Selection," *Journal of Economic Literature*, vol. 34(3), pages 1236-1263

Nguyen, N.X., and F.W. Derrick (1997), "Physicians behavioural response to a Medicare price reduction", Health Services Research 32: 283-298

Pearson S, Schneider E, Kleinman K, Coltin K and Singer J. (2008) "The Impact Of Pay-For-Performance On Health Care Quality In Massachusetts, 2001–2003" *Health Affairs*, 27, no. 4: 1167-1176

Peterson L., Woodard L., Urech T., Daw C., Sookanan S. (2006) "Does pay-for-performance improve the quality of health care". *Annals of Internal Medicine*, 145(4):265-272

Predengast C. (1999) "The provision of incentives in firms" *Journal of Economic Literature*, vol. 37, issue 1, pages 7-63

Richards, J. (2009). "Is there an elephant in the room?" *British Journal of General Practice,* 59: 376-377.

Robinson, J. C. (2001), "Theory and practice in the design of physician payment incentives", Milbank Quarterly, 79(2), 149-77

Rockville, M.D. (2004), "Strategies to Support Quality-based Purchasing: A Review of the Evidence", Agency for Healthcare Research and Quality 2004:04-P024. U.S. Department of Health and Human Services. Public Health Service.

Rosenthal, M. B. and A. Dudley (2007). "Pay-for-Performance: will the latest payment trend improve care?" *Journal of the American Medical Association*, 297 (7): 740-44.

Rosenthal, M. B. and R. G. Frank (2006). "What Is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review* 63(2): 135-157.

Rosenthal MB, Frank RG, Li Z, Epstein AM. (2005) "Early experience with pay-for-performance: from concept to practice." *Journal of the American Medical Association*, 294 (14), 1788-93.

Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. (2006) "Pay for performance in commercial HMOs." *The New England Journal of Medicin*e, 355: 1895-1902

Roski, J., Jeddeloh R., An, L., Lando, H., Hannan, P., Hall, C. and Zhu, SH. (2003) "The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines." *Preventive Medicine,* Volume 36, Issue 3: 291-299

Scott A. (2001) “Eliciting GP’s preferences for pecuniary and non-pecuniary job characteristics.” *Journal of Health Economics*. 20:329-47

Scott, A. and A. Shiell (1997), “Do fee descriptors influence treatment choices in general practice: a multi-level discrete choice model”, Journal of Health Economics 16: 323-342

Shen, Y. (2003) “Selecition incentives in a performance-based contracting system,” *Health Services Research,* 38 (2), 535-552

Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. (2007)"Quality of clinical primary care and targeted incentive payments: an observational study." *British Journal of General Practice*, 2007 June; 57 (539). pp. 449-454.

Stiglitz, Joseph E, (1974). "Incentives and Risk Sharing in Sharecropping," *Review of Economic Studies*, Blackwell Publishing, vol. 41(2), pages 219-55, April

Town R., Wholey D.R., Kralewski J., Dowd B. (2004) “Assessing the influence of incentives on physicians and medical groups.” *Medical Care Research and Review*, 2004 Sep; 61(3 Suppl):80S-118S.

Town R., Kane R., Johnson P., Butler M. “Economic Incenives and physicians’ deliveary of preventive care: a systematic review”. *American Journal of Preventive Medicine*. 2005; 28(2):234-240

Vaghela P, Ashworth M, Schofield P, and Gulliford M. (2009) "Population Intermediate Outcomes of Diabetes under Pay-for-Performance Incentives in England from 2004 to 2008." *Diabetes Care*, March 2009 vol. 32 no. 3 427-429

Wagstaff, A., Moreno-Serra, R. (2009) “Europe and central Asia’s great post-communist social health insurance experiment: Aggregate impacts on health sector outcomes”. *Journal of Health Economics*. 28: 322-340

Wennberg JE, Barnes BA, Zubkoff M. (1982) “Professional uncertainty and the problem of supplier-induced demand.” *Social Science & Medicine*, 16:811-24

Wilson, R. (2006). “Primary Care Renewal in Ontario- Focus on Remuneration”, Presented at the Primary Care Forum, November 26-27, 2006. Airport Marriott, Toronto Ontario. accessible at the following web link-- http://toolkit.cfpc.ca/en/remuneration/appendix-2.php

Wooldridge, J.M., (2002) “Econometric analysis of cross section and panel data.” MIT Press, Cambridge.

Yip, W. (1998), “Physician responses to medical fee reductions: changes in the volume and intensity of supply of Coronary, Artery Bypass Graft (CABG) surgeries in the Medicare private sectors”, Journal of Health Economics 17: 675-700

Young G, Meterko M, Beckman H, Baker E, et al. (2007) "Effects of Paying Physicians Based on their Relative Performance for Quality", *Journal of General Internal Medicine* 2007; Volume 22, Number 6, 872-876

Figure 3.1: Groups of Physicians with Different Timing in PCR Participation and P4P Incentive Exposure

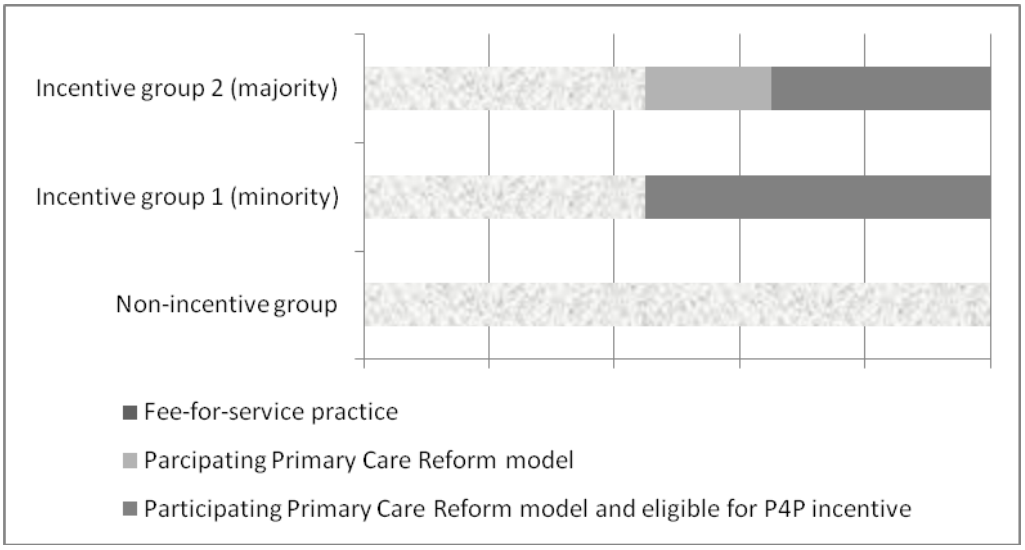


Figure 3.2: Mean Number of Claims for Contact Incentive Payments per Eligible Physician

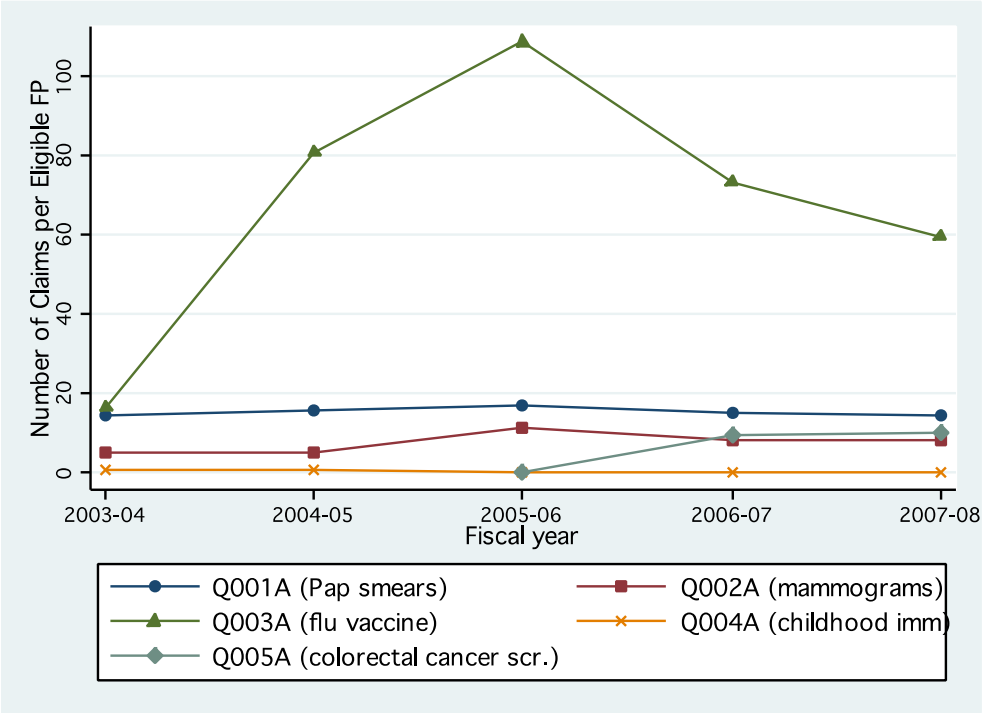


Figure 3.3: Share of Target Practice Population Receiving Targeted Service-- Senior Flu Shot

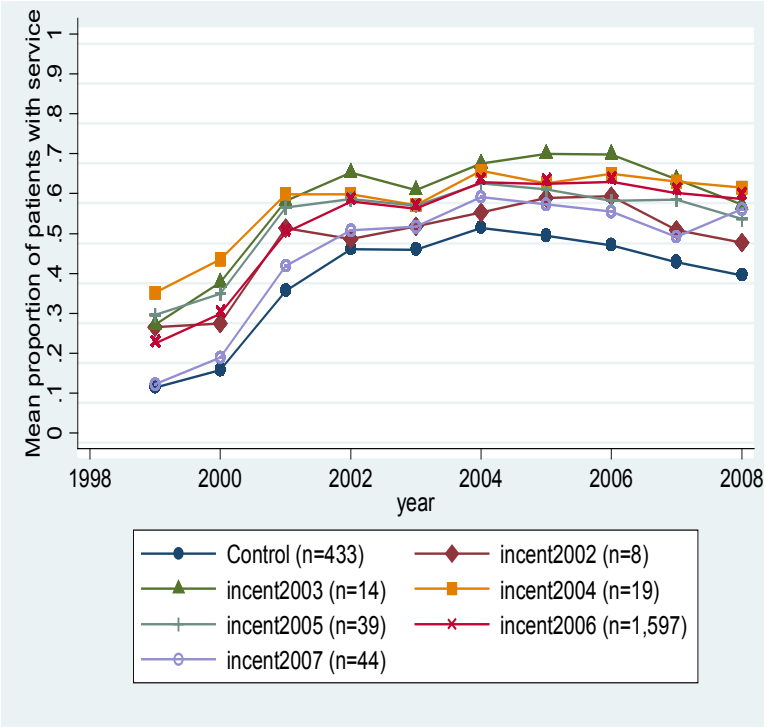


Figure 3.4: Share of Target Practice Population Receiving Targeted Service-- Toddler Immunzation

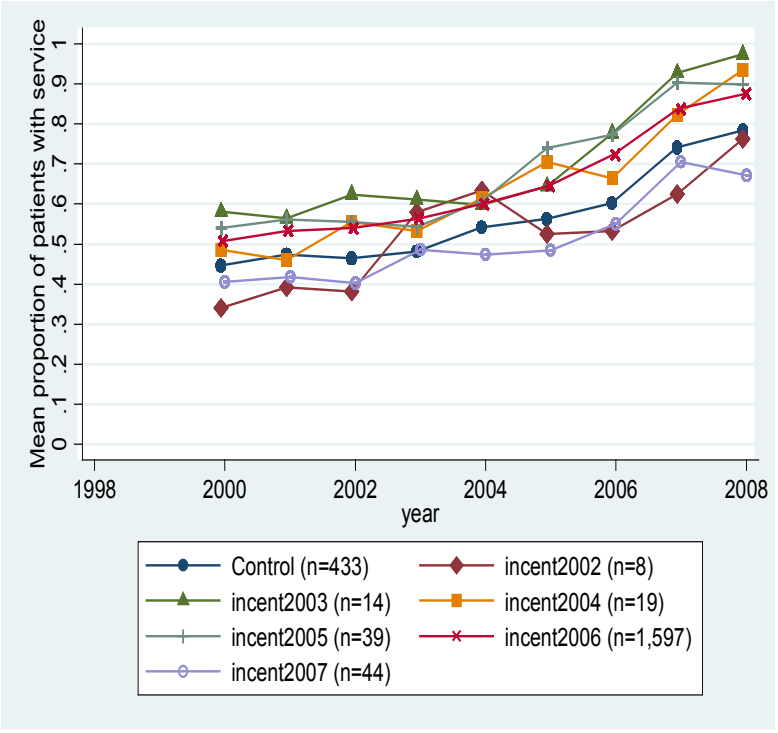


Figure 3.5: Share of Target Practice Population Receiving Targeted Service -- Pap Smear

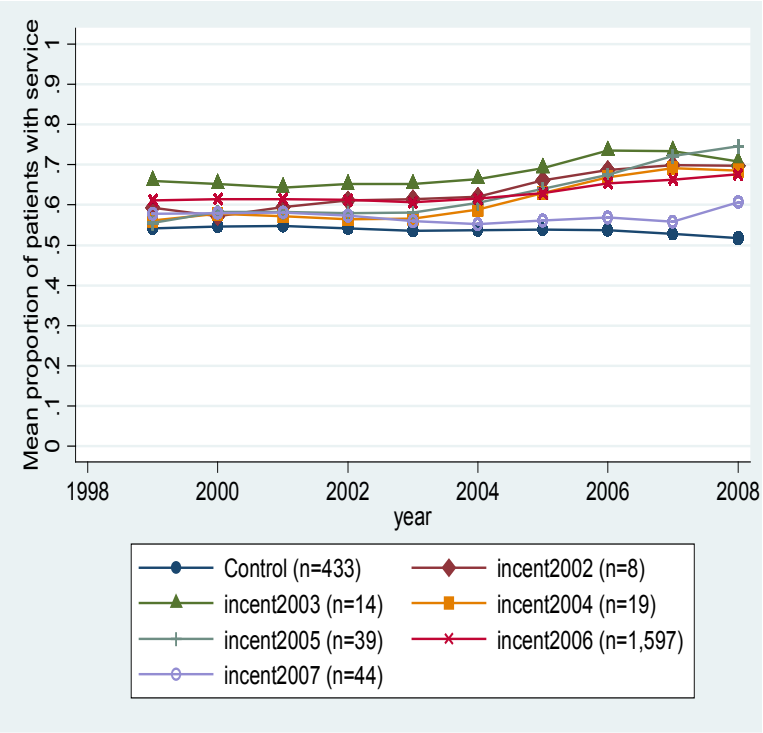


Figure 3.6: Share of Target Practice Population Receiving Targeted Service -- Mammogram

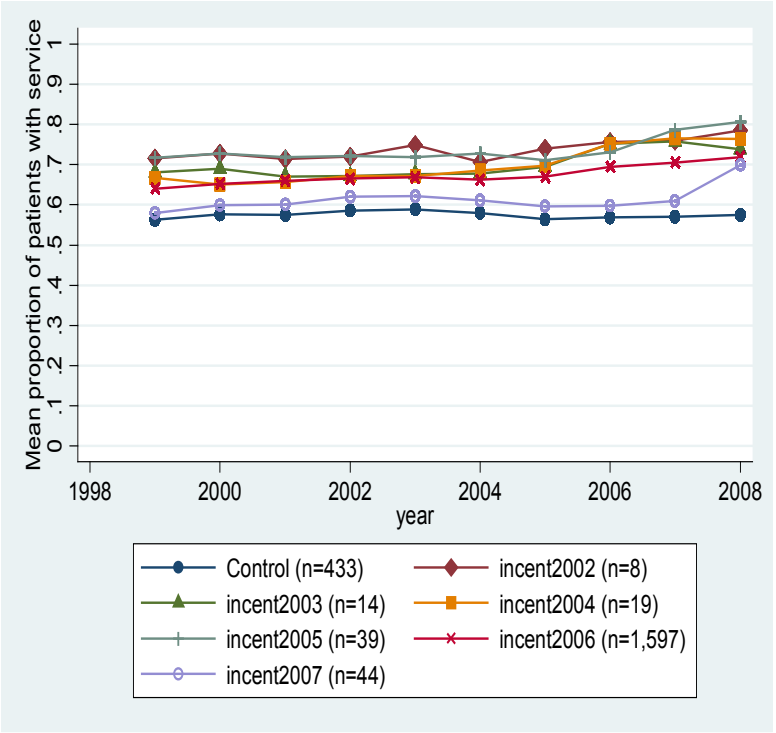


Figure 3.7: Share of Target Practice Population Receiving Targeted Service — Colorectal cancer screening

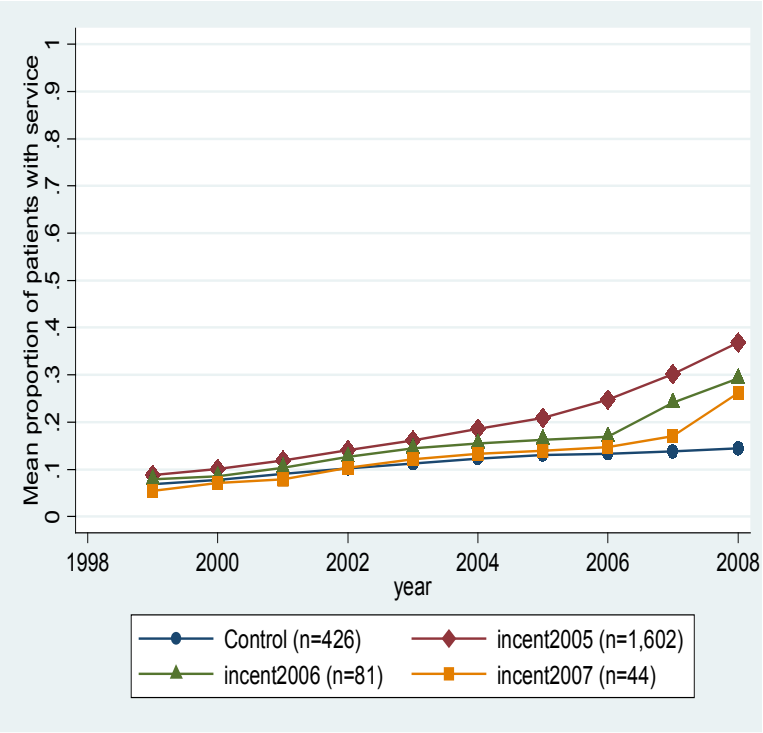


Figure 3.8: Proportion of Physicians Achieving the Targeted Performance Level of Service — Obstetrical deliveries

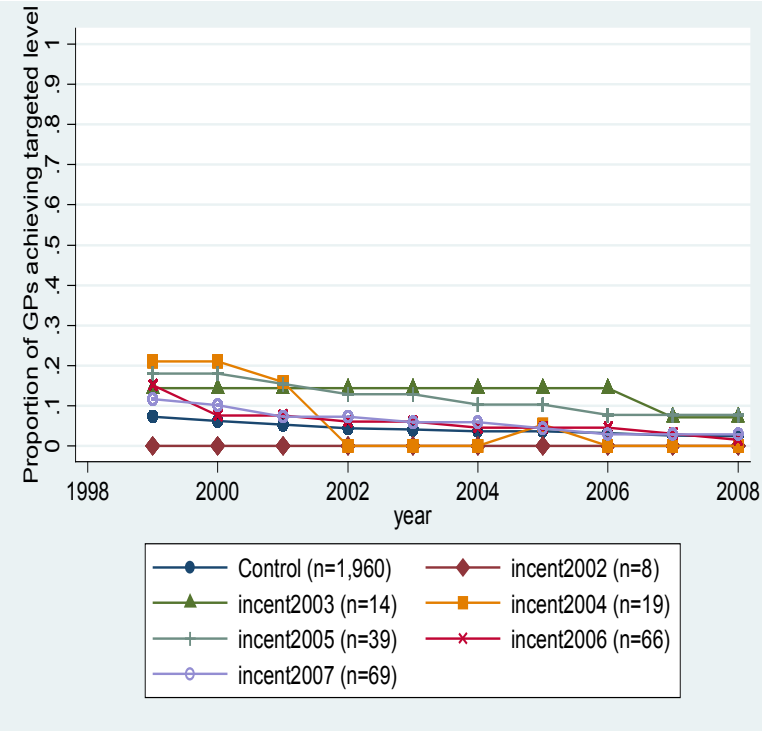


Figure 3.9: Proportion of Physicians Achieving the Targeted Performance Level of Service — Hospital services

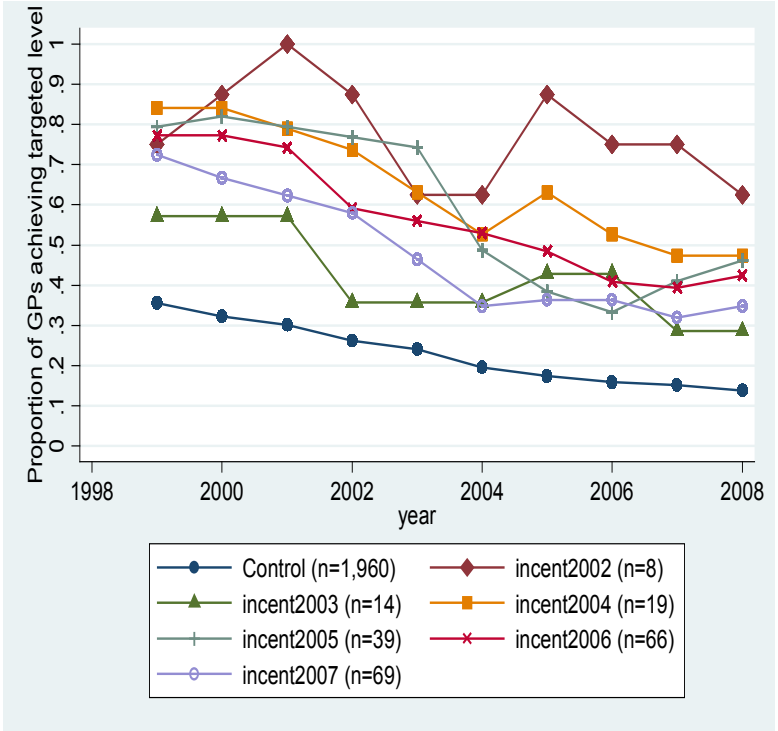


Figure 3.10: Proportion of Physicians Achieving the Targeted Performance Level of Service — Palliative care

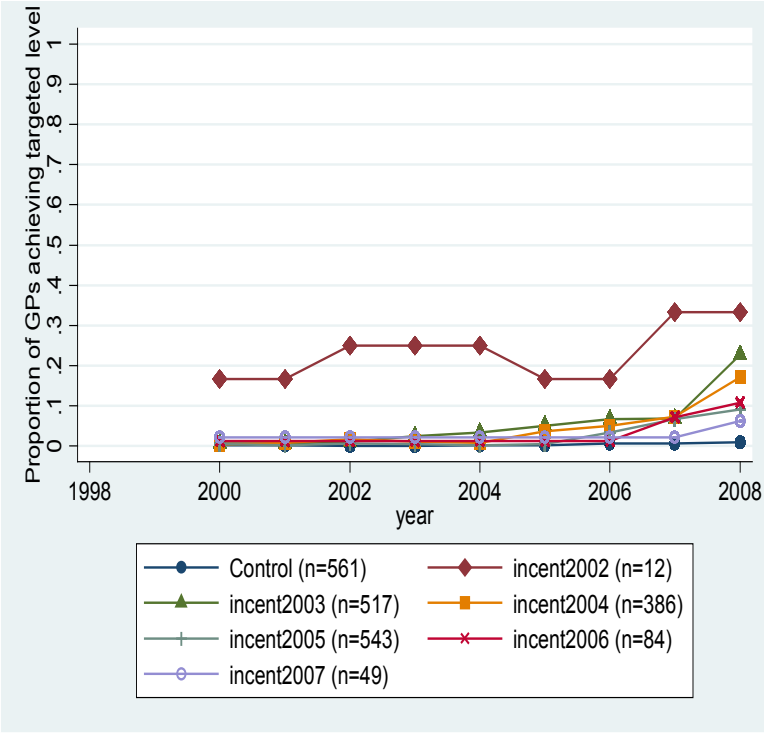


Figure 3.11: Proportion of Physicians Achieving the Targeted Performance Level of Service — Office procedures

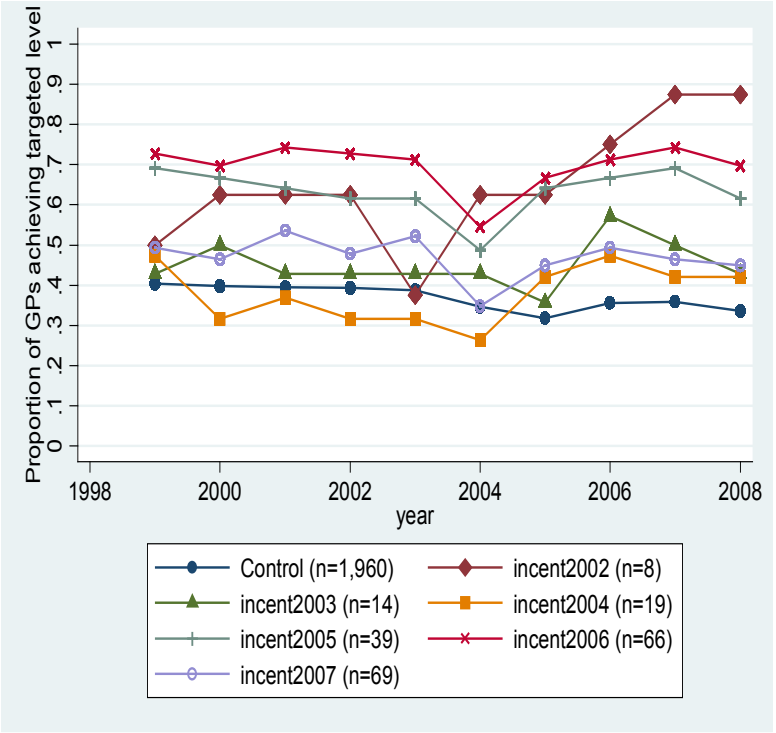


Figure 3.12: Proportion of Physicians Achieving the Targeted Performance Level of Service — Prenatal care

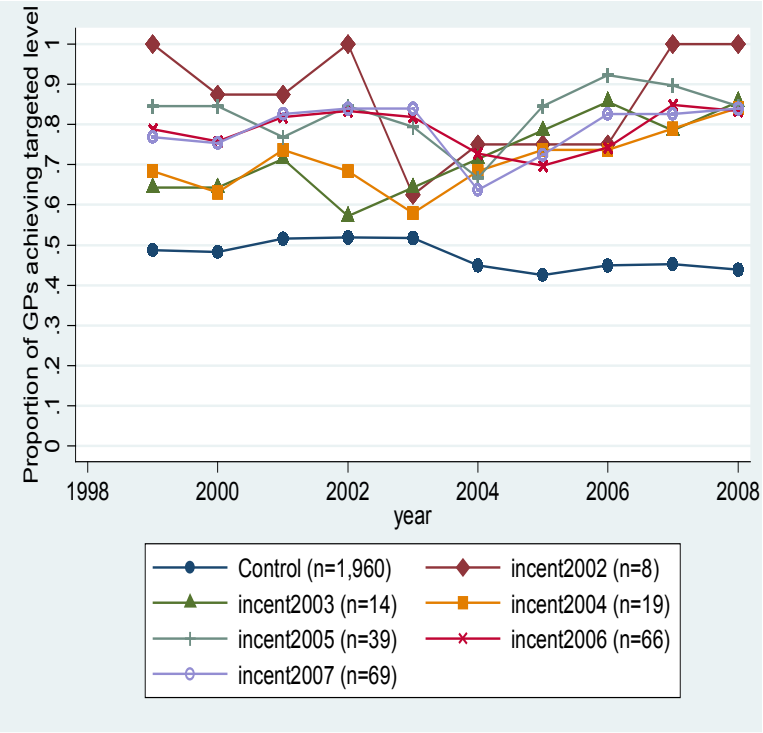


Figure 3.13: Proportion of Physicians Achieving the Targeted Performance Level of Service — Home visits

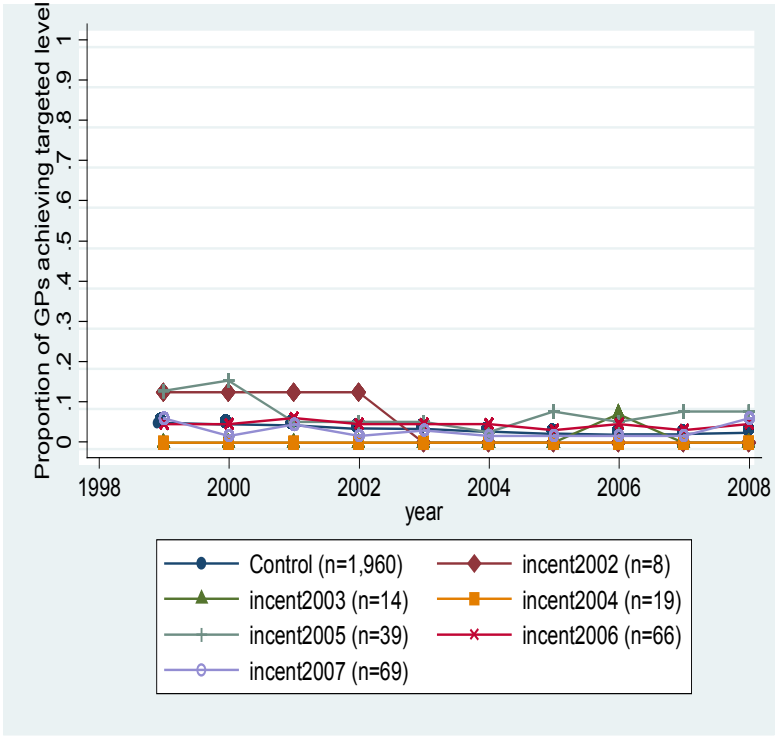


Table 3.1: Description of Eleven Financial Incentives under Analysis

#	Financial incentive	Eligibility condition	Bonus Payment
		Preventive Care Service Enhancement Payments	
Contact Payment			
		-- payment of $6.86 for each documented contact for eligible patients to obtain the preventive service	
Cumulative Care Preventive Service Bonus			
1	Seniors' influenza immunizations	Bonus payment based on the proportion of the physician's eligible (aged 65 or more) and rostered patients on March 31 who received the flu shot in the previous flu season.	• $220 (60% of patients) • $440 (65% of patients) • $770 (70% of patients) • $1,100 (75% of patients) • $2,200 (80% of patients)
2	Pap smears	Bonus payment based on the proportion of the physician's eligible(females aged 35 to 69) and rostered patients on March 31 who received a Pap smear for cervical cancer screening during the last 30 months.	• $220 (60% of patients) • $440 (65% of patients) • $660 (70% of patients) • $1,320 (75% of patients) • $2,200 (80% of patients)
3	Mammograms	Bonus payment based on the proportion of the physician's eligible (females aged 50 to 69) and rostered patients on March 31 who received a mammogram for breast cancer screening during the last 30 months.	• $220 (55% of patients) • $440 (60% of patients) • $770 (65% of patients) • $1,320 (70% of patients) • $2,200 (75% of patients)
4	Toddler immunizations	Bonus payment based on the proportion of the physician's eligible (children aged 30 to 42 months) and rostered patients on March 31 who received 5 immunizations by the age of 30 months.	• $440 (85% of patients) • $1,100 (90% of patients) • $2,200 (95% of patients)
5	Colorectal cancer screenings	Bonus payment based on the proportion of the physician's eligible (aged 50 to 74) and rostered patients on March 31 who was administered a colorectal screening test by Fecal Occult Blood Testing during the last 30 months.	• $220 (15% of patients) • $440 (20% of patients) • $1,100 (40% of patients) • $2,200 (50% of patients)
		Annual Special Payments	
6	Obstetrical deliveries	Payment if 5 or more obstetrical services were delivered to 5 or more patients in a fiscal year.	$3,200 (increased to $5,000 in October 2007)
7	Hospital services	Payment if hospital services provided to all patients total at least $2,000 in a fiscal year.	$5,000 (increased to $7,500 in April 2005 for those with a Rurality Index of Ontario score greater than 45)
8	Palliative care	Payment if palliative care services are delivered to four or more patients in a fiscal year.	$2,000
9	Office procedures	Payment if office procedures provided to enrolled patients total at least $1,200 in a fiscal year.	$2,000
10	Prenatal care	Payment if prenatal care services are provided to five or more enrolled patients in a fiscal year.	$2,000
11	Home visits	Payment if 100 or more home visits are provided to enrolled patients in a fiscal year.	$2,000

Table 3.2: Eligibility for Preventive Care Bonuses and Special Payments

	2002	2003	2004	2005	2006	2007	2008
Preventive Care Bonuses							
Senior Flu Immunization. Toddler Immunization, Pap Smear and Mammogram							
FHN	April						
FHG						April	
CCM						April	
FHO						April	
Colorectal Cancer Screening							
FHN					April		
FHG					April		
CCM					April		
FHO						April	
Special Payments							
Obstetrical Services, Hospital Services, Office Procedures, Prenatal Care and Home Visits							
FHN	April						
FHG				- never eligible -			
CCM				- never eligible -			
FHO					November		
Palliative Care							
FHN	April						
FHG		July					
CCM				- never eligible -			
FHO					November		

Note: Date PCR models introduced: FHN: April 2002; FHG: July 2003; CCM: October 2005; FHO: November 2006.

Table 3.3: Key Characteristics of Primary Care Models Included in This Study

Model (Year Introduced)	**Size/Rostering**	**Funding**
Traditional Fee-for-service (whole period)	No size regulation No rostering	• Fee-for-service
Family Health Network (2002)	*Physician:* • At least 3 GPs *Rostering:* • Minimum total roster of 2400 for group of 3 • Financial penalty for average rosters > 2400 per GP	Blended Capitation: • Age-sex adjusted capitation for rostered patients for 57 core services (about 80% of gross income) • 10% of FFS rate for core services to rostered patients • Access Bonus: 20.65% of the base capitation payment less value of outside use by rostered patients • Monthly comprehensive care capitation payments for formally rostered patients • 100% of FFS rate for core services to non-rostered patients up to $45,000 per physician • 100% of FFS rate for excluded services to either rostered or non-rostered patients
Family Health Group (2003)	*Physician:* • At least 3 GPs *Rostering:* • Voluntary	Blended fee-for-service: • 100% FFS as usual • 10% premium on FFS rate for specified comprehensive care services provided to rostered patients (Ministry-assigned and formally rostered) • Comprehensive care premium • Monthly comprehensive care capitation payments for formally rostered patients • Some premiums/ bonuses paid only for formally rostered patients
Comprehensive Care Model (2005)	*Physician:* • Solo practice *Rostering* • Required • No size regulation	Blended fee-for-service: • 100% FFS as usual • Monthly comprehensive care capitation for rostered patients
Family Health Organization (2006)	*Physician:* • At least 3 GPs *Rostering* • Required • No size regulation	Blended capitation: • Capitation for core services to enrolled patients • Access bonus: maximum of 18.59% of the base rate payment less outside use by rostered patients • 100% FFS for excluded services to all patients and for non-enrolled patients • Monthly comprehensive care capitation payments for formally rostered patients • 100% of FFS rate for core services to non-rostered patients up to $45,000 per physician • 100% of FFS rate for excluded services to either rostered or non-rostered patients

Note: Some of these elements were present for different periods for the different PCR model

Table 3.4: Criteria for Selecting the Sample of GPs for Analysis

Criterion	**Rationale**	**Resulting Sample Size**
All Ontario physicians present in claims data at any point between April 1998 and March 2008		37,422
Exclude physicians not present in all 10 years of the study period	Exclude physicians who began to practice later than April 1998 or who interrupted their practice or who left province during the study period	- 21,415 = 16,007
Exclude physicians whose specialty is not general/family practice during entire study period[a]	Only GPs are eligible for incentives; we exclude those billing as GP while attaining specialization	- 8,533 = 7,474
Exclude physicians who billed less than $30,000 annually	Exclude part-time physicians	- 1,304 = 6,170
Exclude physicians without two consecutive years of practice before start of study (i.e., April 1996 to March 1998)	Exclude new GPs who are newly establishing their practice at the start of the study period	- 3,835 = 2,335
Exclude GPs for which billings for A001, A003, and A007 constitute less than 70% of activity, and GPs for which billings for A001, A003 and A007 constitute less than 50% of all activity and a single "non A-code" category constitutes over 15% of activity[c]	Exclude GP specialists whose main activity is other than providing traditional family medicine visits and consultations[b]	- 95 = 2,240
Exclude locums[d]	Locums are not eligible for the incentives[e]	- 19 = 2,221
Exclude GPs affiliated with the following primary care groups: RNPGA, HSO, PCN, SEAMO, GHC or ICHA[f]	Such GPs did not submit claims data or submitted only shadow billing claims; available data are insufficient for the analysis	- 32 = 2,189
Exclude GPs who converted between FFS and PCR practices for more than one time during the study period	Such GPs do not represent typical observations in service provision behaviour	- 4 = 2,185

[a] A physician's specialty was defined as the specialty under which the largest share of services were billed (based on the fee approved).
[b] An additional criterion whereby a GP with more than 25% of billings for K-codes was classified as a GP psychotherapist was rendered superfluous by this 70% rule.
[c] The following ophthalmology codes are treated as a "non A" category: A009A, A110A, A111A, A112A, A114A, A115A, A237A, A238A, A239A, A240A, E077A.
[d] Locums were identified using information on the Group Type in the Corporate Provider Database. We were able to identify locums only imperfectly.
[e] GPs in walk-in-clinics do not regularly provide the services eligible for the financial incentives under study and should therefore be excluded from the analysis. We considered excluding GPs with a high proportion of billings for code A888 (Emergency Department Equivalent – Partial Assessment); however, once all of the above criteria were applied, this criterion was redundant.
[f] RNPGA: Rural and Northern Physician Group Agreement; HSO: Health Service Organization; PCN: Primary Care Network; SEAMO: Southeastern Ontario Medical Organization; GHC: Group Health Care; IHCA: Inner City Health Associates.

Table 3.5: Definitions and Sample Sizes of Control and Treatment Groups for each Performance Incentive

Preventive Care Incentives		
Senior Flu Shot, Toddler Immunization, Pap Smear, Mammogram		
Control Group	• FFS	433 physicians
Treatment Group	• FHN starting from April 2002 • FHG starting from April 2007 • CCM starting from April 2007 • FHO starting from April 2007	1,722 physicians
Colorectal Cancer Screening		
Control Group	• FFS	427 physicians
Treatment Group	• FHN starting from April 2006 • FHG starting from April 2006 • CCM starting from April 2006 • FHO starting from April 2007	1,730 physicians
Special Payments		
Obstetrical Care, Hospital Services, Office Procedures, Prenatal Care, and Home Visits		
Control Group	• FFS • FHG • CCM	1,962 physicians
Treatment Group	• FHN starting from April 2002 • FHO starting from November 2006	218 physicians
Palliative Care		
Control Group	• FFS • CCM	560 physicians
Treatment Group	• FHN starting from April 2002 • FHG starting from July 2003 • FHO starting from November 2006	1,596 physicians

Table 3.6. Descriptive Statistics in the Pre-intervention Period: Control and Incentive Groups

	Control Group			Incentive Group			Equal Means	Equal Variance
	Mean	Median	St. Dev	Mean	Median	St. Dev	p-value	p-value
Preventive Care Incentives for Senior Flu Shot, Toddler Immunization, Pap Smear, Mammogram, Colorectal Cancer Screening and Special Payment for Palliative Care								
Physician Characteristics								
Age	54.0	54.0	10.8	49.0	49.0	8.4	0.000	0.000
Female	0.213	-	-	0.270	-	-	-	-
Years Licensed	22.1	20.0	11.6	18.1	16.0	8.8	0.000	0.000
Practice Characteristics								
Size	1,408	1,345	668	1,605	1,572	573	0.000	0.000
Patient Age	39.7	39.3	8.5	38.6	38.0	6.2	0.000	0.000
Proportion Female	0.528	0.493	0.121	0.544	0.506	0.111	0.000	0.000
Proportion Infants	0.016	0.014	0.014	0.023	0.021	0.014	0.000	0.009
Proportion Elderly	0.144	0.119	0.111	0.135	0.119	0.079	0.000	0.000
Workload								
Annual Workdays	250.6	249.8	44.2	263.7	261.8	38.8	0.000	0.000
Annual Visits	7,663.0	7,337.8	3,531.0	8,466.2	8,307.5	3066.3	0.000	0.000
Visits/Workday	30.3	29.1	12.8	31.9	31.0	10.2	0.000	0.000
Special Payments for Obstetrical Care, Hospital Services, Office Procedures, Prenatal Care, and Home Visits								
Physician Characteristics								
Age	50.2	50.0	9.3	47.8	47.0	8.1	0.000	0.000
Female	0.255	-	-	0.284	-	-	-	-
Years Licensed	19.1	17.5	9.6	17.8	15.0	8.8	0.000	0.000
Practice Characteristics								
Size	1,571	1,526	608	1,497	1,491	478	0.000	0.000
Patient Age	38.8	38.2	6.9	39.1	38.8	5.3	0.091	0.000
Proportion Female	0.538	0.500	0.114	0.565	0.528	0.102	0.000	0.000
Proportion Infants	0.021	0.019	0.014	0.027	0.025	0.013	0.000	0.000
Proportion Elderly	0.136	0.117	0.088	0.151	0.147	0.069	0.000	0.000
Workload								
Annual Workdays	260.8	260.3	40.4	264.2	261.8	38.9	0.000	0.018
Annual Visits	8,392	8,209	3,230	7,424	7,383	2,415	0.000	0.000
Visits/Workday	31.9	31.0	10.9	27.8	27.8	7.4	0.000	0.000

Notes: the null hypothesis of the t-tests on the equality of means is that the variable has the same mean for the treatment and control groups; the null hypothesis of the F-tests for the homogeneity of variances is that the variable has the same standard deviation for the treatment and control groups.

Table 3.7: Proportion of Eligible Family Physicians who Submitted at Least One Claim for a "Contact Incentive Payment"

	Pap Smear	Mammogram	Senior Flu Vaccine	Toddler Immunization	Colorectal Cancer Screening
2003-04	0.43	0.30	0.22	0.17	-
2004-05	0.28	0.19	0.62	0.13	-
2005-06	0.44	0.36	0.62	0.06	0.02
2006-07	0.37	0.27	0.47	0.06	0.18
2007-08	0.30	0.25	0.40	0.04	0.19

Note: Only physicians in FHNs were eligible from 2003-04 to 2005-06 for the contact incentive payments; FHN and FHO physicians were eligible for 2006-07 and 2007-08; FHGs and CCMs were never eligible during the study period; they became eligible April 1, 2008. FFS physicians were never eligible and remain ineligible.

Table 3.8: Main Results: Preventive Care Bonuses, Estimated Marginal Effects, Difference-in-Differences Models-- Full Sample

		DID with pooled OLS model			DID with physician-specific fixed effects model			DID with differential trend model			Specification test	
	(a)		(b)			(c)			(d)		(e)	
	Baseline Compliance in 2003	Marginal Effect (St. Error)	Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	Sample size: # obs (# GPs)	R^2	Wald Test Statistics	P-value
Senior Flu Shot	0.554	0.013 (0.010)	19,866 (2,029)	0.371	0.028*** (0.007)	19,866 (2,029)	0.470	0.036*** (0.009)	19,866 (2,029)	0.469	2.71	0.100
Toddler Immunization	0.543	-0.007 (0.013)	16,826 (1,999)	0.278	0.011 (0.011)	16,826 (1,999)	0.356	0.004 (0.014)	16,826 (1,999)	0.356	0.66	0.417
Pap Smear	0.589	0.031*** (0.006)	19,926 (2,029)	0.433	0.041*** (0.004)	19,926 (2,029)	0.115	0.050*** (0.006)	19,926 (2,029)	0.115	12.17	0.001
Mammogram	0.646	0.004 (0.007)	19,888 (2,029)	0.351	0.018*** (0.005)	19,888 (2,029)	0.158	0.022*** (0.006)	19,888 (2,029)	0.158	1.44	0.230
Colorectal Cancer Screening	0.150	0.095*** (0.009)	19,918 (2,027)	0.217	0.085*** (0.005)	19,918 (2,027)	0.373	0.113*** (0.006)	19,918 (2,027)	0.379	66.30	0.000

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; * indicates statistical significance at the 10% level.

Table 3.9: Robustness checks: Preventive Care Bonuses, Estimated Marginal Effects, Difference-in-Differences Estimator with Physician-specific Fixed Effects—Alternative Estimation Samples

	GPs who joined PCR before introduction of bonuses as treatment group				GPs in PCR models funded primarily by fee-for-service as treatment group			
	(a) Baseline Compliance in 2003	Marginal Effect (St. Error)	(b) Sample size: # obs (# GPs)	R^2	(c) Baseline Compliance in 2003	Marginal Effect (St. Error)	(d) Sample size: # obs (# GPs)	R^2
Senior Flu Shot	0.561	0.028*** (0.007)	19,073 (1,948)	0.468	0.554	0.024*** (0.007)	18,550 (1,893)	0.471
Toddler Immunization	0.548	0.010 (0.011)	16,162 (1,919)	0.356	0.543	0.010 (0.011)	15,669 (1,863)	0.352
Pap Smear	0.591	0.041*** (0.004)	19,130 (1,948)	0.117	0.589	0.040*** (0.004)	18,607 (1,893)	0.111
Mammogram	0.653	0.017*** (0.005)	19,093 (1,948)	0.152	0.646	0.018*** (0.005)	18,569 (1,893)	0.163
Colorectal Cancer Screening	0.144	0.079*** (0.006)	13,158 (1,341)	0.364	0.150	0.085*** (0.006)	17,778 (1,808)	0.355

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; * indicates statistical significance at the 10% level.

Table 3.10: Sensitivity Analysis 1: Preventive Care Bonuses, Estimated Marginal Effects, Difference-in-Differences Estimator with Physician-specific Fixed Effects—Excluding Q-Codes/Exclusion Codes

		Full Sample			GPs who joined PCR before introduction of bonuses			GPs in PCR models funded primarily by FFS		
	(a) Baseline Compliance in 2003	(b) Marginal Effect (St. Error)	Sample size: # obs (# GPs)	R^2	(c) Marginal Effect (St. Error)	Sample size: # obs (# GPs)	R^2	(d) Marginal Effect (St. Error)	Sample size: # obs (# GPs)	R^2
Senior Flu Shot	0.554	0.013* (0.007)	19,866 (2,029)	0.469	0.013* (0.007)	19,073 (1,948)	0.468	0.011 (0.007)	18,550 (1,893)	0.471
Toddler Immunization	0.543	0.008 (0.011)	16,826 (1,999)	0.352	0.007 (0.011)	16,162 (1,919)	0.352	0.007 (0.011)	15,669 (1,863)	0.349
Pap Smear	0.589	0.024*** (0.004)	19,926 (2,029)	0.084	0.024*** (0.004)	19,130 (1,948)	0.085	0.024*** (0.004)	18,607 (1,893)	0.084
Mammogram	0.653	0.017*** (0.005)	19,888 (2,029)	0.162	0.017*** (0.005)	19,093 (1,948)	0.156	0.017*** (0.005)	18,569 (1,893)	0.167
Colorectal Cancer Screening	0.150	0.068*** (0.005)	19,918 (2,027)	0.341	0.061*** (0.005)	13,158 (1,341)	0.334	0.067*** (0.006)	17,778 (1,808)	0.325

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; * indicates statistical significance at the 10% level.

Table 3.11: Sensitivity analysis 2: Preventive Care Bonuses, Estimated Marginal Effects, Difference-in-Differences Estimator with Physician-specific Fixed Effects—Adding in Non-rostered Patients

		Full Sample			GPs who joined PCR before introduction of bonuses			GPs in PCR models funded primarily by FFS		
	(a) Baseline Compliance in 2003	Marginal Effect (St. Error)	(b) Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	(c) Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	(d) Sample size: # obs (# GPs)	R^2
Senior Flu Shot	0.554	0.043*** (0.006)	20,207 (2,029)	0.466	0.044*** (0.007)	19,398 (1,948)	0.465	0.041*** (0.007)	18,847 (1,893)	0.468
Toddler Immunization	0.543	0.007 (0.010)	17,416 (1,999)	0.372	0.006 (0.010)	16,732 (1,919)	0.372	0.007 (0.010)	16,200 (1,863)	0.367
Pap Smear	0.589	0.042*** (0.004)	20,245 (2,029)	0.099	0.043*** (0.004)	19,435 (1,948)	0.102	0.041*** (0.004)	18,885 (1,893)	0.094
Mammogram	0.646	0.027*** (0.004)	20,228 (2,029)	0.150	0.028*** (0.004)	19,419 (1,948)	0.147	0.027*** (0.004)	18,868 (1,893)	0.153
Colorectal Cancer Screening	0.150	0.089*** (0.005)	20,217 (2,027)	0.365	0.085*** (0.005)	13,364 (1,341)	0.365	0.089*** (0.005)	18,029 (1,808)	0.344

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; * indicates statistical significance at the 10% level.

Table 3.12: Sensitivity analysis 3: Preventive Care Bonuses, Estimated Marginal Effects, Difference-in-Differences Estimator with Physician-specific Fixed Effects—Only FHGs and CCMs Achived Minimum Roster Size

		Full Sample			GPs who joined PCR before introduction of bonuses			GPs in PCR models funded primarily by FFS		
	(a) Baseline Compliance in 2003	Marginal Effect (St. Error)	(b) Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	(c) Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	(d) Sample size: # obs (# GPs)	R^2
Senior Flu Shot	0.560	0.031*** (0.007)	18,156 (1,843)	0.469	0.030*** (0.007)	17,706 (1,798)	0.469	0.028*** (0.007)	16,956 (1,720)	0.471
Toddler Immunization	0.548	0.010 (0.011)	15,480 (1,814)	0.370	0.011 (0.011)	15,094 (1,770)	0.370	0.009 (0.011)	14,423 (1,691)	0.366
Pap Smear	0.591	0.043*** (0.004)	18,212 (1,843)	0.120	0.043*** (0.004)	17,762 (1,798)	0.120	0.043*** (0.004)	17,010 (1,720)	0.116
Mammogram	0.649	0.029*** (0.005)	18,179 (1,843)	0.183	0.028*** (0.005)	17,729 (1,798)	0.181	0.029*** (0.005)	16,977 (1,720)	0.188
Colorectal Cancer Screening	0.154	0.091*** (0.006)	17,877 (1,806)	0.380	0.083*** (0.006)	12,443 (1,262)	0.367	0.090*** (0.006)	15,918 (1,607)	0.363

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; * indicates statistical significance at the 10% level.

Table 3.13: Sensitivity analysis 4: Preventive Care Bonuses, Estimated Marginal Effects, Difference-in-Differences Estimator with Physician-specific Fixed Effects—Dropping Transition Year/First Year of Incentive Exposure

		Full Sample			GPs who joined PCR before introduction of bonuses			GPs in PCR models funded primarily by FFS		
	(a) Baseline Compliance in 2003	Marginal Effect (St. Error)	(b) Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	(c) Sample size: # obs (# GPs)	R^2	Marginal Effect (St. Error)	(d) Sample size: # obs (# GPs)	R^2
Senior Flu Shot	0.554	0.037*** (0.009)	18,329 (2,029)	0.475	0.036*** (0.009)	17,607 (1,948)	0.473	0.033*** (0.009)	17,136 (1,893)	0.476
Toddler Immunization	0.543	0.002 (0.015)	15,365 (1,999)	0.318	0.001 (0.015)	14,760 (1,919)	0.318	0.001 (0.015)	14,328 (1,863)	0.316
Pap Smear	0.589	0.049*** (0.006)	18,389 (2,029)	0.106	0.050*** (0.006)	17,665 (1,948)	0.107	0.050*** (0.006)	17,193 (1,893)	0.102
Mammogram	0.646	0.035*** (0.006)	18,352 (2,029)	0.173	0.035*** (0.006)	17,628 (1,948)	0.169	0.034*** (0.006)	17,156 (1,893)	0.177
Colorectal Cancer Screening	0.150	0.111*** (0.006)	18,381 (2,027)	0.390	0.105*** (0.007)	12,259 (1,341)	0.377	0.112*** (0.007)	16,457 (1,808)	0.372

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; * indicates statistical significance at the 10% level.

Table 3.14: Main Results: Special Payments, Estimated Marginal Effects, Difference-in-Differences Estimator (No Physician-specific Fixed Effects)

		GPs Eligible for Special Payments in 2005			GPs Eligible for Special Payments in 2006			GPs Eligible for Special Payments in 2007		
	(a) Baseline Compliance in 2003	Marginal Effect (St. Error)	(b) Sample size: # obs (# GPs)	PseudoR$_2$	Marginal Effect (St. Error)	(c) Sample size: # obs (# GPs)	Pseudo R^2	Marginal Effect (St. Error)	(d) Sample size: # obs (# GPs)	PseudoR$_2$
Obstetrical Services	0.043	-0.0004 (0.005)	19,934 (1,998)	0.302	-0.004 (0.004)	20,187 (2,025)	0.308	0.013 (0.024)	20,196 (2,028)	0.302
Hospital Services	0.272	-0.013 (0.035)	19,777 (1,985)	0.481	-0.005 (0.074)	20,052 (2,012)	0.482	-0.019 (0.037)	20,138 (2,021)	0.482
Office Procedures	0.405	0.006 (0.064)	19,897 (1,995)	0.167	0.075 (0.127)	20,175 (2,022)	0.171	-0.141*** (0.053)	20,209 (2,026)	0.165
Prenatal Care	0.544	0.314*** (0.107)	19,857 (1,991)	0.295	0.106 (0.070)	20,109 (2,016)	0.295	0.184 (0.127)	20,151 (2,020)	0.294
Home Visits	0.045	0.007 (0.007)	18,814 (1,893)	0.225	0.003 (0.012)	19,251 (1,934)	0.230	0.084 (0.078)	19,557 (1,961)	0.226
		GPs Eligible for Special Payments in 2003			GPs Eligible for Special Payments in 2004			GPs Eligible for Special Payments in 2005		
Palliative Care	0.011	0.009 (0.012)	9,681 (1,078)	0.305	0.004 (0.005)	8,495 (946)	0.347	0.032 (0.031)	9,928 (1,104)	0.301

*** Indicates statistical significance at the 1% level.

Table 3.15: Estimated Marginal Effects: Preventive Care Bonuses by Physician Age, Practice Size, and Baseline Level of Compliance, Difference-in-Differences Estimator with Physician-specific Fixed Effects

	By Age					Practice Size					Baseline Compliance				
	GP Age	Baseline in 2003	Marg. eff. (St Err)	# obs. (# GPs)	R^2	Size	Baseline in 2003	Marg. eff. (St Err)	# obs.(# GPs)	R^2	Quartile	Baseline in 2003	Marg. eff. (St Err)	# obs. (# GPs)	R^2
Senior Flu Shot	< 40	0.549	0.024 (0.022)	2,320 (237)	0.48	< 1K	0.547	0.025 (0.016)	3,270 (336)	0.44	Q1 (Lowest)	0.304	0.036*** (0.014)	4,887 (503)	0.42
	40-55	0.577	0.049*** (0.010)	10,508 (1,073)	0.47	1K-1.5K	0.572	0.017 (0.012)	5,935 (606)	0.46	Q2 and Q3	0.593	0.027*** (0.010)	9,908 (1,009)	0.51
	> 55	0.537	-0.001 (0.010)	7,038 (719)	0.47	> 1.5K	0.557	0.031*** (0.010)	10,661 (1,087)	0.49	Q4 (Highest)	0.747	0.020 (0.015)	5,071 (517)	0.49
Toddler Immunization	< 40	0.527	0.027 (0.026)	2,044 (234)	0.56	< 1K	0.496	0.014 (0.034)	2,402 (313)	0.28	Q1 (Lowest)	0.217	0.026 (0.025)	3,810 (455)	0.40
	40-55	0.577	-0.010 (0.015)	9,095 (1,065)	0.39	1K-1.5K	0.560	0.004 (0.021)	5,042 (601)	0.32	Q2 and Q3	0.558	0.012 (0.014)	8,795 (1,023)	0.44
	> 55	0.503	0.037* (0.020)	5,687 (700)	0.26	> 1.5K	0.552	0.021 (0.014)	9,382 (1,085)	0.43	Q4 (Highest)	0.838	0.015 (0.024)	4,221 (521)	0.30
Pap Smear	< 40	0.620	0.059*** (0.013)	2,334 (237)	0.20	< 1K	0.630	0.040*** (0.011)	3,289 (335)	0.10	Q1 (Lowest)	0.374	0.024** (0.010)	4,929 (503)	0.18
	40-55	0.612	0.053*** (0.006)	10,503 (1,069)	0.14	1K-1.5K	0.617	0.038*** (0.007)	5,926 (604)	0.12	Q2 and Q3	0.594	0.050*** (0.006)	9,904 (1,009)	0.13
	> 55	0.549	0.019*** (0.007)	7,029 (717)	0.09	> 1.5K	0.564	0.037*** (0.006)	10,651 (1,084)	0.14	Q4 (Highest)	0.800	0.038*** (0.010)	5,033 (511)	0.14
Mammogram	< 40	0.653	0.045*** (0.016)	2,322 (237)	0.26	< 1K	0.682	0.017 (0.011)	3,284 (336)	0.11	Q1 (Lowest)	0.438	0.034*** (0.009)	5,031 (515)	0.25
	40-55	0.671	0.016** (0.007)	10,532 (1,073)	0.16	1K-1.5K	0.673	-0.003 (0.008)	5,937 (606)	0.20	Q2 and Q3	0.672	0.018*** (0.006)	9,873 (1,007)	0.16
	> 55	0.625	0.014*** (0.007)	7,034 (719)	0.16	> 1.5K	0.632	0.028*** (0.006)	10,667 (1,087)	0.19	Q4 (Highest)	0.829	0.014 (0.010)	4,984 (507)	0.17
Colorectal Cancer Screening	< 40	0.187	0.147*** (0.019)	2,197 (224)	0.48	< 1K	0.159	0.068*** (0.015)	3,243 (330)	0.37	Q1 (Lowest)	0.015	0.091*** (0.007)	4,918 (502)	0.41
	40-55	0.164	0.081*** (0.008)	10,744 (1,092)	0.39	1K-1.5K	0.145	0.079*** (0.010)	6,059 (617)	0.38	Q2 and Q3	0.078	0.102*** (0.007)	9,938 (1,012)	0.45
	> 55	0.116	0.067*** (0.008)	6,977 (711)	0.30	> 1.5K	0.150	0.090*** (0.007)	10,616 (1,080)	0.39	Q4 (Highest)	0.428	0.071*** (0.015)	5,062 (513)	0.43

*** Indicates statistical significance at the 1% level; ** indicates statistical significance at the 5% level; and * indicates statistical significance at the 10% level.

3.A1 Appendix: Empirical Studies on Physician Response to P4P incentives

Table 3.A1 Summary of Empirical Studies on Physician Response to P4P incentives

Study (Authors)	Study design	Incentives involved (Form of incentives, targeted services)	Incentive level	Results	Context	Intervention duration	Sample size / scale of the experiment
Grady et al. 1997	RCT, random assignment 3 arms: 20 education and reward; 18 education; 23 control (61 practices in total)	Reward with education Mammography referrals	Physician	No effect	U.S. 61 primary care practices in greater Dayton, Ohio and Springfield, Massachusetts	6 months	95 physicians in total
Kouides et al. 1998	RCT, non-random assignment 2 arms: 27 practices in treatment, 27 practices in control	Bonus Influenza immunization rates	Provider group	Positive effect in immunization rate in a Medicare population	U.S. For Medicare population	4 months	62 physician in treatment, 82 in control
Hillman et al. 1998	RCT, random assignment 2 arms: 26 PC sites intervention; 26 PC sites control	Bonus(based on the part of capitation payment)+ feedback regarding compliance with guidelines; Cancer screening guidelines: mammography, breast cancer, pap smear, colorectal screening	Provider group	No effect in compliance scores	U.S. Medicaid HMO (contract with numerous other health plans)	18 months period	52 primary care practices, relatively small sample
Hillman et al. 1999	RCT, random assignment 3 arms: control, feedback of performance only, feedback+ bonus payment	Bonus(based on the part of capitation payment)+ feedback Pediatric immunization	Provider group	No effect in compliance scores	U.S. Medicaid HMO (contract with numerous other health plans)	18 months period	53 pediatric practices, relatively small sample

Fairbrother et al. 1999	RCT, random assignment 4 arms: 15 doctors in control, 15 feedback, 15 feedback+ bonus, 15 enhanced FFS+ bonus	Bonus with performance feedback Childhood immunization rate	Physician level	Partial effect: only feedback+bonus improved childhood immunization rate, but primarily achieved through better documentation	U.S. A low-income urban population	12 months	60 physicians in total
Fairbrother et al. 2001	RCT, random assignment 3 arms: 24 bonus, 12 FFS, 21 control	Bonus Pediatric immunizations	Physician level	Partial effect: significant increase in coverage levels, but the increase is primarily due to better documentation not to better immunizing practices	U.S. A low-income urban population	16 months	57 physicians in total
Roski et al. 2003	RCT, random assignment 3 arms: 15 clinics control; 15 bonus; 10 bonus + computerized patient registry	Bonus Smoking cessation	Provider group	Partial effect: improved in adherence to guidelines (documentation of smoking status and providing advice to quit), but no effect in quitting rate	U.S.	12 months	40 clinics in total
Amundson et al. 2003	Observational Before-after analysis No control group	Bonus + performance feedback Tobacco cessation	Provider group	Positive effect improve physician compliance with the tobacco treatment guideline	U.S. HealthParterners system in Minneapolis	3 years	20 medical groups

Beaulieu and Horrigan 2005	Observational Before-after analysis with control group Treatment group: 21 primary care doctors contracted with Independent Health in upstate New York; Control group: provider groups in the Pacific Northwest	Performance bonus + offered with diabetic registry and group discussion process Process and outcome measures for diabetic care	Physician level	Partial effect: patients treated by the doctors contracted with the program had improvement on 7 out of 9 measures	U.S. New York.	8 months	21 physicians as treatment group
Rosenthal et al. 2005	Observational Before-after analysis with control Treatment group: 163 provider groups contracted with PacifiCare Health systems in California; Control group: 42 provider groups contracted with PacifiCare in the Pacific Northwest	Bonus in Quality Incentive Program Process measures: cervical cancer screening, mammography, Haemoglobin A1c test	Provider group	Partial effect: improved only in cervical cancer screening, not improved in mammography, haemoglobin A1c test	U.S. A large network HMO, PacifiCare Health System introduced Quality Incentive Program to contracted medical groups in California in March 2002	10 months	163medical groups eligible for the bonus
Doran et al. 2006	Observational Cross-sectional regression, Not Before-after analysis	Performance contracting: performance w.r.t. 146 quality indicator	Physician level	Positive effect: high levels of reported achievement	U.K. Large scale, national level pay-for-performance contract in 2004	1 year	8,105 physicians
Levin-Scherz et al. 2006	Observational Retrospective cohort study using before-after trend comparison Treatment group: health plans participating in PCHI P4P contracts; Comparison group: national and Massachusetts State measures	P4P contracts: bonus based on network performance compared to previously agreed targets Adult diabetes and pediatric asthma HEDIS scores	Network level	Positive effect: improvement compared to state and national level	U.S. Beginning in 2001, a provider network Partners Community HealthCare, Inc (PCHI) and the health plans began P4P contracts with bonus payments.	2001-2003	18-75 health plans in PCHI as treatment group

Campbell et al. 2007	Observational Before-after analysis, No comparison group	Performance contracting: Clinical indicators on coronary heart disease, asthma, type 2 diabetes	Physician practices	Partial effect: improved in asthma, type 2 diabetes measures. Not improved in coronary heart disease measures	U.K. Large scale, national level pay-for-performance scheme for family practice in 2004	1998, 2003, 2005	42 family practices, national representative
Millett et al. 2007	Observational Before-after analysis, No comparison group	Performance contracting: smoking cessation among patients with diabetes Proportion of patients with documented smoking cessation advice, prevalence of smoking among patients with diabetes	Physician practices	Positive effect: increased the provision of support for smokers with diabetes in primary care settings	U.K. Large scale, national level pay-for-performance scheme for family practice in 2004	2003, 2005	36 primary care practices
Steel et al. 2007	Observational Before-after analysis, No comparison group	Performance contracting: quality of care for two common chronic conditions: asthma and hypertension Six quality indicators referred to asthma and hypertension subject to incentive payments	Physician practices	Positive effect: significant increase for the six indicators referred to asthma and hypertension linked to incentive payments	U.K. Large scale, national level pay-for-performance scheme for family practice in 2004	2003, 2005	18 primary care practices
Mandel and Kotagal 2007	Observational Before-after mean comparison, No control group	Pay for performance coupled with additional improvement interventions related to the collaborative. Flu shot percentage, controller medication percentage for children with persistent asthma, written self-management plan percentage.	Primary care practices	Positive effect: The initiative resulted in substantive and sustainable improvement in all measures	U.S. The Physician-Hospital Organization (PHO) affiliated with Cincinnati Children's Hospital Medical Center launched an asthma improvement collaborative in October 2003	2003-2006	44 pediatric practices

Felt-Lisk et al. 2007	Observational Before-after mean comparison with comparison group Treatment group: five Medicaid-focused health plans in California participating the LIRR Collaborative Comparison groups: national and state benchmarks, and two plans that were part of the collaborative but no incentives	Bonus payments for improving the measure A HEDIS measure for "well-baby visits" that requires six visits by age fifteen months	Health plans	Partial effect: only small effect in some of the plans	U.S A collaborative P4P effort among seven Medicaid-focused health plans in California during 2003–2005, known as the Local Initiative Rewarding Results (LIRR) Collaborative Demonstration.	2002 versus 2003–05	7 health plans
Young et al. 2007	Observational Retrospective cohort study using before-after trend comparison Treatment group: physicians participating in the program; Comparison group: national and New York State RIPA scores	Incentive program placing physicians at financial risk to receive rewards based on their performance relative to other physicians in the program. 4 diabetes performance measures	Individual physician	Partial effect: modest effect provider adherence to quality standards for a single measure of diabetes care	U.S. A pay-for-performance program of the Rochester (New York) Individual Practice Association (RIPA) between 2000 and 2001.	1999-2004	334 primary care physicians
Coleman et al. 2007	Observational Before-after regression analysis No comparison group	Performance-based compensation HbA1c testing for diabetes care in a low-income patient population	Physician level	Partial effects: dramatic improvements in rate of patients receiving recommended number of HbA1c tests; no effect on rate of physicians providing first HbA1c test, nor improvement in patient outcomes	U.S. A performance-based provider compensation program implemented in January 2004 at Access Community Health Network (ACCESS), a large system of federally qualified health centers (FQHCs) in Chicago.	2002-2004	46 physicians

Cutler et al. 2007	Observational Before-after mean comparison with comparison group; Treatment group: patients with diabetes who were followed by the P4P program (CDCM); Comparison group: patients followed by routine care group	P4P incentive payments Percent of eligible patients received an LDL-C test and attained LDL-C control	Medical groups	Positive effect: Higher rates of performance in the P4P program	U.S. Chronic disease care management (CDCM) program received by the Mercy Medical Group (MMG), which is a 160-provider, multispecialty medical group and has participated in the California P4P initiative	2003-2004	165 patients in the incentive program, 1,694 patients as control
Lawrence et al. 2008	RCT, random assignment 2 arms: 24 clinics with P4P payments, 25 usual care clinics	Bonus payment intervention based on tobacco quitline referrals; Rates of referrals	Clinics	Positive effect: increased referrals rates compared to usual care clinics	U.S. A P4P program targeting clinician referral to statewide quitline services in Minnesota.	September 2005 to June 2006	24 clinics as treatment, 25 clinics as control
Pearson et al. 2008	Observational Before-after mean comparison with comparison group Treatment group: physician groups that receive incentives Groups Comparison group: groups that were matched with incentivized groups on their baseline performance but that did not subsequently receive any incentive	Multiple P4P programs introduced into physician group contracts during 2001–2003 by the five major commercial health plans operating in Massachusetts; 13 Health Care Employer Data And Information Set (HEDIS) Measures	Physician groups	No significant effect: no distinguishable different trends among the treatment and the matched comparison group.	U.S. Multiple P4P programs introduced into physician group contracts during 2001–2003 by the five major commercial health plans operating in Massachusetts.	2001-2003	154 physician groups in total
Campbell et al. 2009	Observational Before-after interrupted time-series analysis, No control group	Performance contracting: Clinical indicators on coronary heart disease, asthma, type 2 diabetes	Physician practices	Partial effect: by 2005, improvement quality for asthma and diabetes but not for heart disease. By 2007, the rate of improvement had slowed for all three conditions.	U.K. Large scale, national level pay-for-performance scheme for family practice in 2004	1998, 2003, 2005, 2007	42 family practices, national representative

Vaghela et al. 2009	Observational Before-after analysis, No comparison group	Performance contracting 3 measures related to diabetes outcomes	Physician practices	Positive effect: significant increase	U.K. Large scale, national level pay-for-performance scheme for family practice in 2004	2004-2005, 2007-2008	Around 8,423 practices
Lee et al. 2010	Observational Before-after mean comparison with comparison group; Treatment group: patients with diabetes who were enrolled in the P4P program; Comparison group: randomly sampled patients with diabetes who had never joined the P4P program	Financial Incentives for increasing comprehensive follow-up visits for diabetes care Number of essential exams/tests; numbers of diabetes-related physician visits and hospital admissions	Physician level	Positive effect: P4P program for diabetes was associated with a significant increase in regular follow-up visits and evidence-based services, and significantly lower hospitalization costs.	Taiwan A pay-for-performance (P4P) program for diabetes care operated by the Bureau of National Health Insurance (NHI) in Taiwan.	2005-2006	12,499 patients as intervention group; 26,172 patients as comparison group
Gavagan et al. 2010	Observational Before-after mean comparison with comparison group; Treatment group: clinics received incentives; Comparison group: clinics with no incentives	Financial incentive for achieving group targets in preventive care; Cervical cancer screening, mammography, and pediatric immunization	Physician level	No significant effect: no significant effect on performance of preventive care	U.S. In 2002, 11 public community health centers in Houston/Harris County were provided performance incentives on 3 quality indicators in preventive care	2002	6 clinics as treatment group; 5 clinics as comparison group
Chung et al. 2010	Observational Before-after mean and trend comparison; No comparison group	Bonus payment to physicians: based on individual physicians' performance on 15 ambulatory quality measures, with a composite score	Physician level	No significant effect: no evident effect of physician-specific incentives	U.S. In 2007, all primary care physicians at Palo Alto Medical Clinic (PAMC), California participated in the physician incentive program.	2007	179 physicians
Boland et al. 2010	Observational Before-after mean comparison; No comparison group	Bonus payments were to be made if the radiologists met goals. Three radiologist report turnaround times (RTAT) Components	Provider individual level	Positive effect: significant decrease in turnaround time after the program	U.S. Massachusetts General (MGPO) Physicians Organization at the Massachusetts General Hospital (MGH) initiated a hospital wide department specific radiologist PFP initiative.	July 2006–March 2009	81 radiologists, 11 subspecialty divisions

Lester et al. 2010	Observational Before-after trend comparison; No comparison group	Financial incentive related to quality indicator; Screening for cervical cancer, control of hypertension, diabetes control, and screening for diabetic retinopathy	Medical facilities	Positive effect: upward trend when incentives were in place and downward trend when incentives were removed.	U.S. (and U.K.) Four of original financial incentives removed for 35 outpatient facilities owned and operated by Kaiser Permanente Northern California.	1999-2007	35 outpatient medical facilities
Mullen et al. 2010	Observational Before-after regression analysis with control (DID) Treatment group: provider groups contracted with PacifiCare Health systems in California; Control group: provider groups in the Pacific Northwest	Bonus in Quality Incentive Program, Another annual bonus program by the Integrated Healthcare Association (IHA), bonus based on cervical cancer screening, mammography, Haemoglobin A1c test, asthma medication	Provider group	Partial effect: improved only in cervical cancer screening, not improved in mammography, haemoglobin A1c test, asthma medication	U.S. A large network HMO, PacifiCare Health System introduced Quality Incentive Program to contracted medical groups in California in March 2002; One year later, PacifiCare with five other big health plans introduced another larger P4P program by the Integrated Healthcare Association (IHA).	2002-2004	Treatment groups size (77-186) medical groups; Control group size: (7-32) medical groups

3.A2 Appendix: All Data Sources

Table 3.A2 Data Sources and Main Relevant Variables

Data Source	Relevant information
Ontario Health Insurance Program (OHIP) physician claims data	Claim records to calculate the services provided by physicians; Basic provider information; Basic patient information; Fee paid/billed; Patient encrypted health number as linking variable; Provider encrypted number as linking variable
Corporate Provider Database (CPDB) data	Physician demographic variables; Physician practice variables; Physician PCR group participation, effective dates; Provider encrypted number as linking variable
Client Agency Program Enrollment (CAPE) data	Patient member status; Patient roster dates; Patient encrypted health number as linking variable; Provider encrypted number as linking variable
Registered Persons Database (RPDB) data	Demographics of registered persons; Postal code of residence; Patient encrypted health number as linking variable

3.A3 Appendix: Assigning Patients to Primary Care Physicians' Practices

The Hutchison methodology (Hutchison, Hurley, Birch, Lomas, & Stratford-Devai, 1997) was implemented for all FPs in the province, not just those in the analysis sample. This ensured that an individual was assigned to an FP as called for by the algorithm, regardless of whether the FP was included in the analysis (if we focused only on the analysis sample, we would have falsely assigned some patients to sample physicians when the individual's real family physician was not included in the sample). A FP's practice population is defined as:

- All persons for whom the physician billed OHIP for at least one visit (see below for how a visit was defined) during the previous fiscal year; and
- All additional patients for whom the physician billed OHIP for at least one visit in each of the two preceding fiscal years.
- Patients who met these criteria for more than one physician were assigned to the physician who billed for the largest number of visits in the most recent year.
- When an equal number of visits were made to more than one physician in the most recent year, assignment is made to the made to the physician who billed for the most recent visit.

A service is defined as a FP visit if:

The attending physician is a FP and the fee code is one of the following 74 visit codes from the Ontario Schedule of Benefits

Fee Schedule Code	**Description**
A001A	Minor assessment
A003A	General assessment
A004A	General re-assessment
A005A	Consultation
A006A	Repeat consultation
A007A	Intermediate assessment or well baby care
A008A	Mini assessment
A110A	Periodic oculo-visual assessment, aged 19 years and below
A112A	Periodic oculo-visual assessment, aged 65 years and above
A115A	A major eye examination
A888A	Emergency department equivalent – Partial assessment
A901A	House call assessment – First patient seen
A902A	House call assessment – Pronouncement of death in the home
A903A	Pre-dental/operative general assessment (maximum of 2 per 12-month period)

Fee Schedule Code	Description
A905A	Limited consultation
A933A	On-call admission assessment
A945A	Special palliative care consultation
E070A E071A	Geriatric Geriatric Age Premium: Gen. Practice - Geriatic Gen. Assess. Premium - 75 or Older
E075A	Geriatric general assessment premium – patient aged 75 or older (maximum 1 per 12 month period)
E077A	Identification of patient for a Major Eye Examination
G212A	Diagnostic and Therapeutic Procedure – Hyposensitisation, including assessment and supervision – When sole reason for visit
G271A	Diagnostic and Therapeutic Procedure – Cardiovascular – Anticoagulant supervision – long-term, telephone advice
G365A	Diagnostic and Therapeutic Procedure – Papanicolaou Smear – periodic
G372A	Diagnostic and Therapeutic Procedure – Intramuscular, subcutaneous or intradermal – With visit (each injection)
G373A	Diagnostic and Therapeutic Procedure – Intramuscular, subcutaneous or intradermal – Sole reason (first injection)
G538A	Active immunization – Injection of unspecified agent – with visit (each injection)
G539A	Active immunization – Injection of unspecified agent – sole reason (first injection)
G590A	Active Immunization – Injection of influenza agent – With visit
G591A	Active Immunization – Injection of influenza agent – Sole reason
K004A	Family psychotherapy – 2 or more family members in attendance at the same time
K005A	Primary mental health care - Individual care – Per half hour
K006A	Hypnotherapy – Individual care – Per half hour
K007A	Psychotherapy – Individual care – Per half hour
K010A	Psychotherapy – Additional units per member (maximum 6 units per patient per day)
K011A	Hypnotherapy – Group – for induction and training for hypnosis (maximum 8 people), per member, per half hour
K012A	Psychotherapy, Group – Per member of a group of 4, first 12 units per day
K013A	Counselling – Individual care – Per half hour
K017A	Annual health or annual physical examination – Child after second birthday
K019A	Psychotherapy, Group – Per member of a group of 2, first 12 units per day
K020A	Psychotherapy, Group – Per member of a group of 3, first 12 units per day
K022A	HIV Primary Care – Individual care per half hour
K023A	Palliative care support – Individual care per half hour

Fee Schedule Code	Description
K024A	Psychotherapy, Group – Per member of a group of 5, first 12 units per day
K025A	Psychotherapy, Group – Per member of a group of 6 to 12, first 12 units per day
K026A	Family Practice & Practice in General - Certification of medical eligibility for Ontario Hepatitis C Assistance Program
K027A	Family Practice & Practice in General – Certification of medical eligibility for Ontario Hepatitis C Assistance Program (OHCAP)
K028A	Family Practice & Practice in General – Sexually transmitted disease management
K030A	Family Practice & Practice in General – Diabetic management assessment
K031A	Health Protection and Promotion Act – Physician Report – Completion of Physician Report in accordance with Section 22.1 of the Health Protection and Promotion Act.
K033A	Counselling – Individual care – Additional units per patient per provider per 12-month period
K040A	Group counselling – 2 or more persons
K041A	Group counseling – 2 or more persons – Additional units
K070A	Home care application – Application
K071A	Home care supervision – Acute Home Care Supervision (maximum 1 every 2 weeks for the first 12 weeks following admission to home care program).
K072A	Home care supervision – Chronic Home Care Supervision (maximum 1 per month commencing in the 13th week following admission to the home care program).
K623A	Certification of mental illness – Form 1 – Application for psychiatric assessment in accordance with the *Mental Health Act* – includes necessary history, examination, notification of the patient, family and relevant authorities and completion of form.
P004A	Obstetrics, prenatal care – Minor prenatal assessment
W001A	Non-emergency long-term care in-patient services – Subsequent visits – Chronic care or convalescent hospital – additional subsequent visits (maximum 4 per patient per month)
W002A	Non-emergency long-term care in-patient services – Subsequent visits – Chronic care or convalescent hospital – first 4 subsequent visits per patient per month
W003A	Non-emergency long-term care in-patient services – Subsequent visits – Nursing home or home for the aged – first 2 subsequent visits per patient per month
W004A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – General re-assessment of patient in nursing home
W008A	Non-emergency long-term care in-patient services – Subsequent visits – Nursing home or home for the aged – additional subsequent visits (maximum 2 per patient per month)
W102A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Admission assessment – Type 1

Fee Schedule Code	Description
W104A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Admission assessment – Type 2
W105A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Consultation
W106A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Repeat Consultation
W109A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Admission assessment – Annual physical examination
W121A	Non-emergency long-term care in-patient services – Subsequent visits – Nursing home or home for the aged – Additional visits due to intercurrent illness
W107A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services, Admission assessment – Type 3
W777A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Intermediate assessment - Pronouncement of death
W872A	Non-emergency long-term care in-patient services – Subsequent visits – Nursing home or home for the aged – Palliative care
W882A	Non-emergency long-term care in-patient services – Subsequent visits – Chronic care or convalescent hospital – Palliative care
W903A	Emergency or Out-Patient Department (OPD) & Non-Emergency Long-Term Care In-Patient Services – Pre-dental/pre-operative general assessment (maximum of 2 per 12-month period)

3.A4 Appendix: Independent Variable Specification

Table 3.A3 Variable Names and Descriptions

Variable Name	Variable Description
treatment	Treatment dummy indicating whether the GP was eligible for the financial incentive on that date.
post	Pre- & post- dummy indicating whether it is post-intervention period
treated	Treated & untreated dummy indicating whether the this physician was entitled to collect bonus/special payment
ffsd	PCR dummies indicating the physician was affiliated with the model on that date
doc_age	Age of physician in years on March 31 2002
doc_ageg1 -	Physician age-- four categories: doc_age1: physician age < 45; doc_age2: 45≤ physician age < 50; doc_age3:
doc_male	Physician sex: 1 if male; 0 if female
cmaca	Practice location$_{S:}$ metropolitan area influence measured by Metropolitan Influence Zone (five categories: 0 if
riotype	Practice location$_{S:}$ urban/rural level measured by RIO score (Rurality Index of Ontario)
yrslicyr	Years since licensing year as of March 31 2008
lnbsbill	Log of billings: log of total value (fee approved) of claims submitted in 1998-99
workdays	Days of working: total number of days billed in this fiscal year
sum_visit	Number of visits: total # of patient visits in this year
sum_visitg1 -	Total # of patient visits in this year: four categories. sum_visitg1: < 5,000; sum_visitg2: 5,000≤ x < 7,500;
visitpd	Number of patient visits per working day: the number of visits divided by number of days worked in this year
sum_pt	Practice size: number of assigned patients seen per year
sum_ptg1 -	Practice size: number of assigned patients seen per year, four categories
meanage	Average practice age: average age of eligible patient population on each snapshot date
fmpercent	Proportion of females in the practice: proportion of female patients as of eligible patient population on each
clpercent	Proportion of children patients: Proportion of the eligible patients that are under 2 years of age
elpercent	Proportion of elderly patients: Proportion of the eligible patients that are over 65 years of age
year	Year fixed effects:10 dummy variables for each snapshot year from March 31 1999 to March 2008
cdname	Geographical fixed effects: 49 dummy variables defined by Census Division codes

Conclusion

This thesis has empirically investigated a range of important questions related to the economics of health and health care. Health is an important form of human capital that influences social and economic success over the life cycle. Health development during pre-adulthood is critical for future health, educational and economic outcomes over the lifespan. Two of the thesis chapters explored the effects of important factors such as family socio-economic status, environmental characteristics and early life-time events in determining the health development process and the dynamics of health during the period of childhood to early adulthood. The first chapter focused on the roles of family socio-economic conditions and neighbourhood environments in determining the child physical health development, and the second chapter explored the effects of family socio-economic conditions and childhood stressful life-events on the distribution of youth depression, which is one of the most common mental health conditions during adolescence. The third chapter of the thesis focused on another important theme in health economics-- physician behaviours, and investigated whether and

how a certain type of financial incentives--- pay-for-performance payments--- could motivate physician behaviours of service provisions to improve the quality of health care.

Each thesis chapter is a self-contained piece of research, and each makes contributions to the current literature in these areas of health economics. Both the first and the second chapters provide results that shed light on the understanding of health human capital development during pre-adulthood, and also provide important implications for the design of public policies that aim to improve health outcomes of young people. The first chapter contributes to the child health literature as few studies in the child health literature have been focused on modeling the evolution process of health outcomes from childhood to adolescence, particularly in Canada. Moreover, because this paper used information on both family social economic status (SES) and neighbourhood level characteristics in the dynamic panel data framework, it contributes to the health dynamics literature by examining the impact of contextual factors.

The results from the dynamic models indicate strong positive state dependence of child health over time, and the persistence level differs systematically across different neighbourhood status, including average household income, education, and lone-parents status. Results also indicate that it is the

long-term neighbourhood or environmental conditions, rather than short-term variation in conditions, that are contributing to the differences in child health transitions. The predictions from the analyses based on long-term neighbourhood status indicate that children living in poorer neighbourhoods and in neighbourhoods with lower education levels tend to experience poor health status for longer periods, while children tend to experience multiple health drops living in poorer neighbourhoods, in neighbourhoods with less educated people, in neighbourhoods with more families headed by lone-parents and in neighbourhoods with more families living in rental accommodations. These results highlight the importance of targeting the social support programs or social policies to the children who live in these types of disadvantaged communities.

The second chapter examined the roles of family SES, early childhood life-events and unobserved heterogeneity in explaining the distribution of depression among adolescents and young adults. It also explicitly modeled the depression dynamics and quantifies the persistence of depression from adolescence to early adulthood. One contribution of the chapter is that it tackled the methodological issues in modeling the distribution and the dynamics of depression outcomes. First, this chapter employed a conditional quantile regression framework to explore potential heterogeneity in the effects of these

factors across different quantiles of the depression score. Compared with conditional mean estimation models, which have been widely used in this literature, this approach allows the examination of the differential effects of the factors of interest at different parts of the youth depression distribution, therefore providing a more complete view of the links between these factors and youth depression. Using a conditional quantile approach also provided an opportunity to explore the source of discrepancies found in the existing empirical literature. Another methodological contribution of this chapter is that, in addition to standard dynamic quantile regression models, it employed a newly developed instrumental variable quantile regression for dynamic panel with fixed effects model to examine the dynamics of depression.

The results from the static models are in line with the majority of the literature, which highlights the important roles of gender, race, birth order, maternal drinking and smoking behaviour during pregnancy, and a set of family SES factors including maternal education and family poverty status. More importantly, the results reveal the asymmetry of the link between stressful life events and youth depression, which is masked by conditional mean estimator. This might explain why some studies observe the adverse effect of stressful life events on youth depression while others do not. Specifically, different types of

life-events have different roles across different quantiles of the depression score: the incidence of family problems (family member loss or parental divorce) during childhood plays a more important role at the higher end of the depression distribution, while the incidence of emotional trauma during childhood plays a more important role at the lower end. The family SES-youth depression gradient varies substantially across different quantiles of the depression distribution: maternal education, maternal unemployment duration and family poverty status are more important at the higher ends of the depression distribution, and are statistically insignificant at the lowest quantile of the depression distribution. These results provides important policy implications of devoting resources to individuals with the most severe level of depression and employing policies that aim to improve these specific family SES conditions for them. The results from the dynamic models show the importance of taking into account individual heterogeneity when examining the dynamics of youth depression. The pooled model suggests that there is a strong positive state dependence of youth depression across all quantiles of the CES-D distribution, and the magnitude of the state dependence estimate is larger at higher ends of the depression distribution, indicating a higher persistence level for the individuals who have more severe depressive symptoms. However, results from the instrumental

variable with fixed effects model deliver a different message: the pure state dependence of youth depression is very low and the observed positive association between previous depression and current depression is mainly due to unobserved individual heterogeneity.

The first two thesis chapters provide some important messages about the dynamics of health during childhood and adolescence. First, it is important to account for unobserved individual heterogeneity, and to disentangle pure state dependence from unobserved heterogeneity when modeling health dynamics during this period. In both papers, the perceived state dependence in health is substantially attributable to unobserved individual heterogeneity. Second, while child physical health exhibits a certain degree of persistence over time conditional on individual heterogeneity, indicating the permanent nature of physical health accumulation, adolescent depression appears to be mean-revert (again conditional on individual heterogeneity). Third, health production in pre-adulthood is largely associated with contextual effects that are at both the family and community levels. Both papers stressed the important roles of family social economic factors, particularly the mother's role (e.g. maternal education and maternal employment) in the process of child health capital formation.

The third thesis chapter exploited a natural experiment in the province of

Ontario, Canada to identify empirically the impact of pay-for-performance (P4P) incentives on the provision of targeted primary care services. The main contribution of this study is that it provides direct policy implications related to the employment of P4P incentives to improve health care quality, which is extensively employed and strongly advocated in recent years by health plans and governments in many countries. The overall small physician responses to the introduction of P4P incentives in Ontario indicate the rather low power of using these incentives to motivate high quality care. Several lessons can be learned from this study. First, since the cost of complying may vary substantially among different types of procedures and services, tailoring the absolute size of financial incentives for different targeted services according to the relative costs of complying may provide a more cost-effective solution. Second, the findings suggest that there is only limited scope for using P4P incentives to increase the provision of targeted services, and the employment of P4P incentives is only effective when the targeted performance or tasks are strongly linked to professional standards of high quality care. Therefore, future implementations of P4P incentives could be restricted only to these services. Finally, the P4P incentives should be redesigned so that the target measures are more closely related to real standards of high quality care. For example, financial incentives can

be linked to quality indicators that aim to increase access to health care, or to those representative of evidence-based health care.

The third chapter also contributes to the literature on empirical examination of physician response to the pay-for-performance incentives, because it exploited the natural experiment of the primary care reform in Ontario, which provides a good setting for this type of study. The timing of the reforms allows the employment of a difference-in-differences approach to control for potential sources of selection bias and confounding. Another advantage of this study is that it used a unique administrative data set that provides rich information for the investigation of physician service provisions.

This thesis reveals some important areas for future research and extensions. Two of the thesis chapters underscored the impact of family-level environmental factors. One direction for future studies is to attempt to better identify the causal effect of these factors. Identifying the causal effects of family SES faces common empirical challenges due to potential reverse causation and omitted variables that may be correlated with both family SES and child health outcomes. Therefore, future studies should employ good instruments that generate exogenous variation in the family SES factors to study the causal effects, and to explicitly examine the mechanisms through which these family attributes lead to heterogeneous child

health outcomes. Another direction for future studies is to study the dynamic features of the health accumulation process during pre-adulthood. The thesis indicates the importance of accounting for the dynamic aspect of child health. Moreover, the recent skill formation literature has established a life-cycle investment framework to study the formulation of cognitive and non-cognitive skills, which also reveals the dynamic feature of the human capital accumulation process. The literature documents the "dynamic complementarities" and "self-productivity" features of the cognitive and non-cognitive skills, and indicates that these features may as well exist for child health. Therefore, more research can be done to examine these features in the child health accumulation process using longitudinal data sets and advanced dynamic panel data models.

On the other hand, the third thesis chapter delivered a general message about physician behaviour: physicians do not automatically respond to performance-based financial incentives as expected. Although principal-agent theory suggests the potential of employing P4P incentives to motivate physicians to provide high-quality care, physician responses to such incentives are not easily predicted. The heterogeneity of physician responses found in the study suggests that physician behaviours may be constrained by a complex set of objectives that we do not directly observe. Therefore, more refined positive analyses on

physician labour supply and service provision behaviours are warranted to inform future implementation of incentive schemes or public policies to elicit desired physician behaviours.