

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Economics

2012

Empirical Studies in the Economics of Health Insurance, Health, and Fertility

Geruso, Michael L.

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01p5547r417>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

EMPIRICAL STUDIES IN THE ECONOMICS OF HEALTH
INSURANCE, HEALTH, AND FERTILITY

Michael L. Geruso

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF ECONOMICS

Adviser: Anne C. Case

June 2012

© Copyright by Michael L. Geruso, 2012.

All rights reserved.

Abstract

This thesis is a collection of essays in the economics of health insurance, health, and fertility. The three essays each explore distinct and unconnected research questions. The first study asks whether differences in health risk can explain the systematically different health plan choices made by younger and older US consumers. Finding the answer to be no, the essay then explores the implications of these differential selection patterns by age. The second study asks how much of the glaring disparities in life expectancy between blacks and whites in the US can be accounted for by group differences in socioeconomic factors. The third study asks whether the often-observed negative correlation between educational attainment and fertility reflects a causal relationship. The first two essays are single-authored, with the second published in the journal *Demography*. The third essay is co-authored with Damon Clark and Heather Royer.

These three essays share an emphasis on empirical rather than theoretical contributions, and in all cases they benefit from access to unique or restricted datasets. Nonetheless, the essays differ in the ways they characterize the data used, and the extent to which the quantities of interest are non-parametrically identified. The empirical evidence in the health insurance essay relies most heavily on economic theory for identification. In part, this first essay attempts to identify deep parameters associated with expected utility theory. The fertility essay, in contrast, relies on a natural experiment to identify its parameters of interest, which are the behavioral responses to an educational reform, rather than primitives of a utility function. Finally, the essay on racial disparities in mortality pushes aside the issue of identification, and instead provides a novel and careful accounting of partial correlations between race, socioeconomic status, and mortality. These correlations are important because they have not been previously reported, despite much interest in the research question.

The first essay, in more detail, examines how health plan choice varies by age and health risk, and asks whether current pricing restrictions in employer-sponsored health insurance are optimal. US employers cannot charge different insurance premiums on the basis of

age, gender, or other observable characteristics. Using administrative employer data, I show that younger and older employees in the same firm have very different preferences over health plans, above and beyond what differences in their health risk would predict. I build a simple model of plan choice that demonstrates that as long as this is the case, requiring that young and old employees face the same prices can result in socially inefficient self-sorting across the various plans offered by the employer. The phenomenon is distinct from the well-known adverse selection effects. I estimate a structural model of plan choice to quantify the efficiency gains that would result if pricing were allowed to vary with age. The results indicate that the welfare gains from introducing age-adjusted prices are small, though they are similar in magnitude to recent estimates of the welfare gains from perfectly correcting adverse selection under uniform prices in employer health plans.

The second essay quantifies the extent to which socioeconomic and demographic characteristics can account for black-white disparities in life expectancy in the United States. Although many studies have investigated the linkages between race, socioeconomic status, and mortality, this article is the first to measure how much of the life expectancy gap remains after differences in mortality are purged of the compositional differences in socioeconomic characteristics between blacks and whites. The decomposition is facilitated by a reweighting technique that creates counterfactual estimation samples in which the distribution of income, education, employment and occupation, marital status, and other theoretically relevant variables among blacks is made to match the distribution of these variables among whites. For males, 80% of the black-white gap in life expectancy at age 1 can be accounted for by differences in characteristics. For females, 70% percent of the gap can be accounted for.

The third essay examines the effect of school attendance and educational attainment on fertility. To shed new light on the causal relationship, we exploit a natural experiment generated by a change in UK compulsory schooling laws in 1972, which raised the minimum leaving age from 15 to 16. The reform was recent enough that access to legal abortion and

modern contraception at the time was quite similar to today, granting unique insight into the fertility effects of education in a modern context. We show that the affected girls had significantly lower fertility in their teen years and that the decline was not accompanied by an increase in abortions. We also find that the reform had a negligible impact on completed fertility. Our findings suggest that education-based policies might reduce teen pregnancies without impacting completed fertility rates.

Acknowledgements

I am sincerely grateful to my advisers Anne C. Case and David Lee for their consistent and generous guidance during my graduate study. I would also like to express my gratitude to Harvey S. Rosen and Tom Vogl, who served on my committee. Several other people were influential in the creation of this dissertation, and more generally in my development as an economist. They include Orley Ashenfelter, Damon Clark, Angus Deaton, Thomas Espenshade, Henry Farber, Marco Gonzalez-Navarro, Bo Honoré, Zhuan Pei, Uwe Reinhardt, Andrew Robinson, Sam Schulhofer-Wohl, and Dean Spears. John Papp deserves special mention. Participants at the Princeton Labor workshop, the Princeton Public Finance Working Group, and the Princeton Center for Health and Wellbeing workshop provided many helpful suggestions. Princeton's Center for Health and Wellbeing and the Industrial Relations Section provided financial support. Finally, I am deeply thankful for the love and encouragement of my family—especially April and Bob Geruso, whose confidence in me never wavered.

Contents

Abstract	iii
Acknowledgements	vi
1 Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences	1
1.1 Introduction	1
1.2 Sources of Inefficiency	6
1.3 Data Source and Measurement of Risk	12
1.3.1 Data: Administrative Health Claims	12
1.3.2 Risk Measurement	15
1.4 Reduced Form Results	17
1.4.1 Demand Patterns Across Age and Gender	17
1.4.2 A Sufficient Statistic for Expected Costs	20
1.4.3 What Drives Demand Differences?	23
1.5 Structural Estimates	25
1.5.1 Discrete Choice Model	25
1.5.2 Expected Costs	28
1.5.3 Identification	29
1.5.4 Estimation Details	31
1.5.5 Parameter Estimates	33

1.6	Policy Simulations	34
1.6.1	Welfare Metric	34
1.6.2	Counterfactual Pricing	35
1.6.3	Welfare Analysis	37
1.7	Conclusion	38
2	Black-White Disparities in Life Expectancy: How Much Can the Standard SES Variables Explain?	54
2.1	Introduction	54
2.2	Data	58
2.3	Methods	61
2.3.1	Decomposition Strategy	61
2.3.2	Generating Reweighting Functions and Constructing Counterfactuals	62
2.3.3	Relationship to Other Demographic Decompositions	65
2.4	Results	67
2.4.1	Main Results	67
2.4.2	Robustness	73
2.5	Discussion	74
2.6	Conclusion	76
3	The Impact of Education on Fertility: Quasi-Experimental Evidence from the UK	87
3.1	Introduction	87
3.2	Compulsory Schooling Laws and Fertility Trends	91
3.2.1	Compulsory Schooling Laws	91
3.2.2	Fertility Trends	93
3.3	Data	95
3.4	Empirical Strategy	98

3.5	Results	100
3.5.1	Education	100
3.5.2	Teen Fertility	102
3.5.3	Cumulative Fertility	104
3.6	Discussion	106
3.6.1	Teen Fertility	107
3.6.2	Completed Fertility	109
3.7	Conclusions	111

List of Tables

1.1	Characteristics of enrolled employees	40
1.2	PPO cost-sharing rules for individuals, 2005-2007	41
1.3	Health plan selection and utilization	41
1.4	Regressions of plan choice on health risk	42
1.5	Fit of ACG prediction	43
1.6	Parameters from expected utility plan choice model	44
1.7	Comparisons of the estimated risk coefficient	45
1.8	Pricing and welfare changes under counterfactual pricing policies	46
2.1	Summary statistics at baseline	78
2.2	Comparability between the NLMS and NCHS	79
2.3	Decomposition of e_1 : Results for males, reweighting the black sample . . .	80
2.4	Decomposition of e_1 : Results for females, reweighting the black sample . .	81
2.5	Decomposition of e_1 : Results for males, reweighting white sample	82
2.6	Decomposition of e_1 : Results for females, reweighting white sample	83
2.7	Decomposition of e_{30} : Results for males, reweighting the black sample . . .	84
2.8	Decomposition of e_{30} : Results for females, reweighting the black sample . .	85
2.9	Sensitivity analysis: Alternative categorizations of income	86
3.1	Data sources used in the study	113
3.2	Effects of ROSLA on age of exit from school	114
3.3	Effects of ROSLA on teen fertility	115

3.4	Effects of ROSLA on cumulative fertility	116
3.5	Effects of ROSLA on husband's qualifications	117
3.6	Age 17 fertility: Robustness to alternative specifications	118
3.7	Total fertility: Robustness to alternative specifications	119

List of Figures

1.1	Equilibrium and Efficient Insurance Allocations	47
1.2	Community Pricing with Heterogeneity in Willingness to Pay	48
1.3	Local Linear Regressions of Plan Choice Versus Healthcare Consumption	49
1.4	Local Linear Regressions of Plan Choice Excluding Drugs	50
1.5	Local Linear Regressions of Plan Choice by Gender	51
1.6	Local Linear Regressions of Plan Choice Versus Last Period's Consumption	52
1.7	Local Linear Regressions of Plan Choice Versus Ex-Post Consumption	53
3.1	Impact of ROSLA on Educational Attainment	120
3.2	Historical Total Fertility Rates in the UK, US, France, and Germany	121
3.3	Historical Teen Fertility Rates in the UK, US, France, and Germany	122
3.4	Impact of ROSLA on Average Age of School Exit	123
3.5	Impact of ROSLA on Fertility at Age 16	124
3.6	Impact of ROSLA on Fertility at Age 17	125
3.7	Impact of ROSLA on Fertility at Ages 18 and 19	126
3.8	Impact of ROSLA on Cumulative Fertility to Age 45	127
3.9	Evolution of Cumulative Effects from Ages 16 to 45	128
3.10	Detailed Timing of Teen Birth Effects	129
3.11	Impact of ROSLA on Teen Abortions	130
3.12	Effects on Husband's Educational Qualifications	131
3.13	Effects of ROSLA on Age-Specific Fertility: Ages 16 to 45	132

3.14 Cross-Validation: ROSLA's Effect on TFR	133
--	-----

Chapter 1

Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences

1.1 Introduction

Health insurance markets in the US are heavily and increasingly regulated. Some interventions are intended to correct classic problems of adverse selection, while others are meant to shape markets to conform to various notions of fairness. Employer health plans, the most common source of health insurance in the US, are governed by federal rules prohibiting employers from charging different premiums or offering different coverage options to their employees on the basis of characteristics like age, gender, or medical history.¹ Employers must offer their health plans at common (but arbitrary) prices to all of their similarly-situated employees. This prohibition against price discrimination, which occurs in a variety of insurance markets to various degrees, is usually motivated by a concern for equity, but it is widely acknowledged that achieving equity via uniform prices can come at the cost of allocative efficiency (e.g. Cutler and Reber 1998; Carlin and Town 2008; Bundorf et al. 2008; Einav and Finkelstein 2011). This is because even without restric-

¹The Employee Retirement Income Security Act of 1974 (ERISA), amended by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), maintains extensive rules governing employee benefit plans. Just as in employment itself, these regulations rule out distinctions in health benefits coverage made on the basis of an employee or dependent's gender, race, age, national origin, religion, or disability. They additionally bar health benefit plans from discrimination on the basis of health status or genetic information. These regulations are limited to employer settings only, and do not apply to privately-purchased insurance, though some state laws impose various pricing restrictions in private health insurance markets.

tions on pricing, consumers with the highest privately-known costs to insure will select the more comprehensive insurance contracts. This drives up the price of more comprehensive insurance and drives out the lower-risk consumers (Akerlof 1970). Restricting price discrimination has the potential to exacerbate this adverse selection problem because it rules out offering a lower price to even *observably* lower-risk agents.²

Much empirical work in insurance markets generally, and in health insurance in particular, has focused on searching for evidence of this kind of adverse selection on risk. This paper examines selection along other dimensions of private information and its implications for efficient sorting into contracts under a uniform pricing constraint. Most of our standard intuition about selection in insurance markets relies on the assumption that risk is the only important determinant of insurance demand and that people with the same insurable risk have the same willingness-to-pay for insurance. But individuals facing the same risk might value insurance differently, for instance, because of differences in risk aversion. Recent empirical work in several insurance markets has indeed shown that willingness-to-pay for insurance can deviate significantly and systematically from underlying risk (e.g. Finkelstein and McGarry 2006; Cohen and Einav 2007; Fang et al. 2008). Cutler et al. (2008) provide a survey of this literature and suggest that, “heterogeneity in preferences may be as, or more, important than heterogeneity in risk in explaining insurance demand.” How this heterogeneity might interact with various pricing regulations has not been well-explored.

In this paper, I ask first whether different groups of consumers (men and women; older and younger individuals) select across plans differently, conditional their expected costs to insure. After finding that they do, I explore whether this preference heterogeneity can interact with uniform pricing restrictions in a way that creates welfare losses beyond those due to the classic adverse selection problem. In other words, even if adverse selection price distortions were addressed via known remedies, such as subsidizing the adversely selected

²For instance, Buchmueller and DiNardo (2002) found that following the introduction of community rating in the individual health insurance markets in New York, healthier consumers were pushed into HMOs due to rising prices in more flexible health plans.

contract, could the rules that ban price discrimination between, say, consumers of different ages still introduce inefficiencies that could only be avoided under separate pricing?

In employer health plans in particular, there is a special reason to look beyond the concern that adverse selection creates inefficiently high prices: In employer plans, this is solvable in principle. Absent preference heterogeneity, adverse selection poses an efficiency problem only because of a feedback mechanism from average costs in the risk pool back into prices. But employers are not constrained to link prices to average costs; they heavily subsidize their plans and can adjust their subsidies to achieve arbitrary (uniform) prices for each plan. In particular, they can set subsidies to exactly compensate for the plan price distortions created by adverse selection (Cutler and Reber 1998).³

To explore the interaction between pricing regulations and demand heterogeneity, I first extend a standard model of insurance selection to highlight the basic theoretical inefficiency arising from pricing regulations when the willingness-to-pay for insurance differs across groups. I show that if insurance valuation, conditional on objective risk, is correlated with some observable characteristic, then ruling out price discrimination on the basis of that characteristic can reduce efficiency. The intuition is straightforward: Efficient prices are determined by the intersection of the willingness-to-pay and marginal cost curves. But observably different groups may have systematically different willingness-to-pay, even conditional on their expected costs of insuring. Therefore, inducing efficient sorting via prices requires setting different prices for different groups.⁴ My model is closely related to the theoretical frameworks of Bundorf et al. (2008) and Einav and Finkelstein (2011). Both of those papers note that in the face of demand heterogeneity, even optimal uniform prices are only constrained efficient, and that first-best pricing would require price discrimination on

³This remedy to adverse selection loss is special to subsidized settings and contrasts with a simple competitive setting in which equilibrium prices equal the average costs of the respective enrollee pools. Both types of markets are discussed in the model in Section 2.

⁴This result is special to markets with asymmetric information. In a typical goods market, a social planner could achieve an efficient allocation simply by setting price equal to marginal cost. Such a solution is not possible in insurance markets because cost is a function of who purchases the contract, rather than a function of the seller's production technology.

the basis of privately-known expected costs. In contrast, the insight and focus of the model here is that welfare-improving price discrimination on the basis of observables is possible, even when the observables are signals only of willingness-to-pay and not of expected costs.⁵

Next, using detailed administrative health claims from a single large employer, I show that there is indeed significant variation in willingness-to-pay for insurance that is correlated with age and independent of health risk. For instance, 50-59 year-old workers in my employer dataset are 50% more likely than 18-29 year-old workers to choose more comprehensive insurance, conditional on enrolling no dependents and having the same expected costs to insure. These results contribute to a growing body of evidence on the prevalence of preference heterogeneity in insurance markets. The fact that demographic groups differ not only in underlying risk, but also in insurance demand conditional on risk, is a pattern necessarily missed by the many studies that have, in the absence of other risk measures, used demographic variables as proxies for health risk. Here, expected and actual medical costs are precisely measured based on line-item medical billing records from a panel of employees.

Interpreting the reduced form results in light of the model suggests that welfare losses due to the interaction of demand heterogeneity and uniform pricing are likely, but does not reveal the size of the effect. As noted in Einav et al. (2010a), the magnitudes of observed selection effects give almost no insight as to the size of the welfare losses they induce. To quantify these welfare losses, I set up a structural model of plan selection that characterizes the choice between insurance contracts as a choice between money lotteries. I estimate the parameters of a utility function via simulated maximum likelihood. With these parameter estimates, I can run policy simulations to measure the efficiency loss associated with typical employer-subsidized premiums relative to various counterfactual pricing schemes. First, I

⁵Bundorf et al. (2008) note: “If consumers cannot be priced on the basis of their tastes, and tastes matter sufficiently, achieving efficient sorting requires measuring and pricing health risk.” Here I explore the case in which consumers can, indeed, be priced according to tastes—at least to the extent that tastes are correlated with observable characteristics.

follow the recent literature to estimate gains from switching to optimal Cutler and Reber (1998) premiums, still subject to the constraint of no price discrimination. These ‘optimal’ premiums solve the adverse selection distortion via an appropriate employer subsidy to the price of the adversely-selected plan. (If preference heterogeneity were unimportant, then these prices would lead to efficient insurance allocations.) The resulting gains are around \$11 per enrollee per year, which are very similar to estimates in Carlin and Town (2008) and Einav et al. (2010a).

In the second counterfactual, I set optimal prices separately by age group. Consistent with my extension to the standard model, I find that optimal age-specific prices differ from the optimal uniform price, with the older enrollees facing a larger contribution for more comprehensive insurance. Also consistent with my model, I find that the age-adjusted prices yield a welfare improvement beyond the optimal uniform prices. The total gain is \$12 per enrollee per year. In a third counterfactual, I find the welfare gain of applying typical employer subsidy rules separately to each age group. This counterfactual is relevant because the informational requirements needed to implement optimal prices are stringent. I find that most (95%) of the welfare gain from switching from uniform prices based on typical employer subsidies to uniform prices based on optimal employer subsidies can be achieved by applying the same typical subsidy rules separately to each age group.

The results indicate potential gains from allowing price discrimination on the basis of age in this setting. However, the dollar value of these gains is small, even though the optimal age-specific prices differ significantly from each other. An important finding from the recent empirical literature is that suboptimal pricing in health insurance need not lead to meaningful welfare losses. My findings echo that notion, with the implication that in employer-sponsored plans, the restriction against price discrimination in the interest of equity may carry relatively small efficiency costs, even once one allows for demand heterogeneity. Nonetheless, these efficiency calculations ignore zero-sum transfers, and uniform pricing involves, *prima facie*, large implicit transfers from younger to older workers.

These implicit transfers are eliminated under age-adjusted pricing.

The remainder of the paper is organized as follows. Section 2 explains intuitively why prohibiting price discrimination can cause an inefficiency when willingness-to-pay for insurance is correlated with observable characteristics independently of risk, and that this inefficiency cannot be corrected by even an omniscient employer (planner) free to set any uniform price. Section 3 describes the data, and Section 4 presents a set of reduced form empirical results that demonstrate that willingness-to-pay for insurance varies with age, conditional on the risk people face. Section 5 presents a structural model of plan choice that builds on the basic reduced-form findings. The structural estimates are used to calculate the welfare changes under alternative pricing policies in Section 6. The conclusion discusses broader implications of the findings.

1.2 Sources of Inefficiency

Here, I provide an intuition for why uniform pricing can be inefficient relative to categorical (separate) pricing on observables, as long as demand varies with these observables in a way that is at least partially uncorrelated with costs to insure. I begin with a standard model of selection, to which I add demand heterogeneity. The model follows in the spirit of Akerlof (1970) in that contract prices, but not the features of contracts, respond endogenously to the selection patterns of consumers. This is in line with most of the recent literature (see, for example, Einav and Finkelstein 2011), and I believe it is particularly relevant for the employer-sponsored health plan setting.

Consider risk averse agents facing a choice between two health insurance options, a high coverage plan (H) and a low coverage plan (L). All agents must choose one of the two plans. The restriction to exactly two options both simplifies the graphical analysis and parallels the menu of options in the empirical exercise that follows, but the restriction is otherwise unimportant. In the empirical exercise, plan L corresponds to a plan with less consumption smoothing across health shocks, but it could alternatively represent the

option of selecting no insurance.

As a starting point, let agents differ only in their health risk, r . Health risk here may be “private” information either in the sense that risk is unobservable or in the sense that contracts cannot be priced on observable correlates of risk because of regulations. Mandated uniform pricing (also referred to as “community rating” in the individual markets) simply expands the set of effective private information to its logical extreme, requiring all agents to face the same price.

Efficient sorting into insurance contracts depends on an agent’s valuation of the contracts relative to her costs to insure. Denote the agent’s difference in valuation between plans H and L as $\Delta WTP(r)$, for willingness-to-pay. It is a function of r because people with different health risk will value health insurance differently. Denote the difference in expected costs of enrolling a type- r in plan H rather than L as $\Delta MC(r)$. $\Delta MC(r)$ reflects differences in the insurer’s expected payouts to medical service providers between the two plans. These payouts can differ between plans because the plans cover different services and/or because they shift different shares of the total bill to the consumer. Efficient allocation requires that $\Delta WTP(r) \geq \Delta MC(r)$ for all agents that choose plan H, and $\Delta WTP(r) < \Delta MC(r)$ for all agents that choose plan L. In other words, the efficiency condition is that an agent chooses the fuller coverage contract if and only if her valuation of the contract exceeds the insurer’s cost of insuring her.⁶

In contrast to the efficiency condition, which relates valuation to cost, the agent’s choice problem relates valuation to price. The agent chooses plan H if $\Delta WTP(r)$ exceeds the price difference between plans, P . A social planner setting insurance prices to maximize welfare subject to the constraint of uniform pricing would face the following problem:

$$\max_P \int \mathbf{1}[\Delta WTP(r) > P] \cdot (\Delta WTP(r) - \Delta MC(r)) dF(r). \quad (1.1)$$

⁶I ignore any externalities from healthcare consumption.

Figure 1.1 displays the efficient allocation that solves this problem, and also shows why both competitive and employer-subsidized prices are unlikely to achieve it. The graphical analysis closely follows Cutler and Reber (1998) and Feldman and Dowd (2000). New features of the analysis are not added until Figure 1.2 below. Health risk is drawn along the horizontal axis, with higher expected costs to the right. Willingness-to-pay is increasing in health risk, so that sicker employees attach a higher relative valuation to plan H. $\Delta MC(r)$ is also assumed to be increasing in r , indicating that as health worsens, the insurer's costs of enrolling the agent grow faster in the more generous plan. Note that average costs, $\Delta AC(r)$, are a function of r only in the sense that the marginal r -type selecting plan H determines the composition of the risk pools in L and H in equilibrium.⁷ As drawn in the figure, $\Delta MC(r)$ and $\Delta WTP(r)$ exhibit single crossing at point D, leading to the efficient price P^* and cutoff risk level r^* . By the efficiency condition, agents with $r < r^*$ should choose plan L, and those with $r \geq r^*$ should choose plan H. Prices set to P^* induce this efficient allocation.

Under competitive pricing, the equilibrium occurs at point E, defined by the intersection of the willingness-to-pay and the *average* cost curves, since the simple competitive equilibrium results in plan prices equal to average plan costs.⁸ At point E, too few agents are insured in in Plan H, the classic adverse selection problem.

In employer health plans, the prices that employees face depend on the employer's scheme for contributing to plan costs.⁹ Therefore, employer prices may be higher or lower

⁷The average cost curve lies above the upward-sloping marginal cost curve because enrollment in plan H increases moving from right to left, not left to right. To transform the diagram so that it accords with the intuition that an average cost curve should lie above a downward-sloping marginal cost curve, simply flip the horizontal axis.

⁸In a competitive pricing environment with no loads and an interior solution, a zero profit condition requires that the price difference would simply be equal to the full difference in average costs between plans, following the $\Delta AC(r)$ curve.

⁹Following Cutler and Reber (1998), a general expression for the employer's contribution to the plan premium is $\alpha_j + \beta \cdot AC_j$, where α_j is a fixed dollar contribution to plan j , β is a percent subsidy, and AC_j is the average cost of providing plan j , which is dependent on the composition of enrollment. It is common for employers to choose either $\alpha_j = A$, $\beta = 0$ which corresponds to a voucher that can be used on a plan of the employee's choice, or $\alpha_j = 0$, $\beta < 1$ which corresponds to an equal percentage subsidy to all plans.

than the efficient prices depending on how much the employer subsidizes premiums.

Cutler and Reber (1998) note that pricing in employer plans is unlikely to be efficient given typical employer contribution practices, but also that it is possible to set an efficient price as long as the employers, or their hired insurance administrators, have some knowledge of the demand and cost curves. The efficient pricing rule simply sets the price employees face in moving from plan L to H equal to P^* . This solution has been noted elsewhere as well (e.g. Feldman and Dowd 2000; Pauly and Herring 2000). Employer plans are especially suited to achieving efficient pricing in this way because employee contributions to plan premiums can be manipulated arbitrarily (as long as they are uniform), and the price-setting employer potentially has an interest in achieving an efficient outcome.

So far I have assumed—in line with most of the adverse selection literature—that insurance demand is solely a function of health risk. However, recent empirical work on selection in insurance markets has shown that willingness-to-pay for insurance can deviate significantly and systematically from underlying risk, and may be correlated with other consumer characteristics. Preference heterogeneity has been found to be of first-order importance in long term care insurance (Finkelstein and McGarry 2006), life insurance (Cutler et al. 2008), Medigap coverage (Fang et al. 2008), and auto insurance (Cohen and Einav 2007).

I now add heterogeneity to the standard model to show that if there are systematic preferences for insurance that are partially uncorrelated with risk but correlated with observables, it is not in fact possible to set a single price for insurance that achieves an allocation that is efficient relative to the optimal categorical prices based on observables. In Figure 1.2, I extend the case in Figure 1.1 to a setting in which two groups, A and B, have different willingness-to-pay for plan H as a function of health risk. The marginal cost curve depends only on risk, not preferences, and so is necessarily identical for the two groups.

Group B has a higher relative valuation of plan H at all risk levels. A number of

phenomena could generate this pattern, including differences in income or wealth, greater risk aversion among group B, or a mean-preserving spread of health risk among individuals in B relative to A. For the purpose of the present analysis, the existence of differential willingness-to-pay, rather than its source, is the salient feature.

In Figure 1.2, the intersections of the marginal cost curve and the willingness-to-pay curves generate two separate efficient prices for the two groups, P_A^* and P_B^* , and two separate efficient risk cutoffs. From an efficiency standpoint, Group A members should choose plan H if and only if their risk exceeds r_A^* . Group B members should choose plan H if and only if their risk exceeds r_B^* . No single price can simultaneously achieve efficient allocations in both groups.¹⁰ The bottom panel in the figure highlights this result. In the bottom panel, Group A achieves an efficient allocation at the pictured common price, but Group B does not. Group B members with risk types between r_B^* and r_B^e should be sorted into plan H rather than plan L because they value plan H above their costs of enrolling in it. For the example in Figure 1.2, even a social planner with full information and arbitrary price setting power could not induce allocations through uniform pricing that are efficient relative to optimal group-specific prices. This is because no single horizontal price line could intersect both point D_A and point D_B .

If the members of groups A and B were unidentifiable in practice, such as if willingness-to-pay were completely idiosyncratic and uncorrelated with observables, then there would be little hope of achieving a more efficient allocation than what is possible under uniform pricing. For instance, Bundorf et al. (2008) make the point that first-best allocations in

¹⁰A numerical example may be helpful. Consider 4 people inhabiting the same community-rated insurance pool. The {WTP, Cost} pairs for persons 1 through 4 are {19, 18}, {15, 16}, {9, 8}, and {5, 6}, respectively. It is efficient only for persons 1 and 3 to take up insurance, since only their willingness-to-pay exceeds their costs of insuring. Note that no single community price can induce this allocation: Prices must be > 15 to keep out person 2, but must also be < 9 to induce in person 3. This is not the typical adverse selection problem because it does not arise solely from the feedback between the cost of providing insurance and the price charged. (Indeed, assume that prices can be set arbitrarily.) Call persons 1 and 2 “old” and 3 and 4 “young.” If young and old can be priced separately, even competitive (average cost) pricing creates efficient allocations, with the competitive price among old being set at 18, and among the young at 8.

selection markets with preference heterogeneity require prices based on private information, with the implication that feasible pricing is generally inefficient. However, the point here is that if these demand differences are correlated with observable characteristics such as age or gender, then a welfare improvement is possible by pricing separately on the basis of these characteristics.

Note that the finding that groups with different willingness-to-pay should optimally face different prices is special to insurance markets (or selection markets, more generally), because unlike with other goods, the cost of production depends on who purchases the good rather than on some production technology of the supplier. In a typical goods market, a social planner could always achieve efficient allocations by setting price equal to marginal cost. It is often argued that this close link between the demand and cost curves is the defining characteristic of insurance markets (see, for example, Einav and Finkelstein 2011).

Note that the inefficiency on display in Figure 1.2 has nothing to do with differences in health risk between the two groups—no assumptions were made on the distribution of health risk within in each group. Instead, it is due to differences in demand, conditional on health risk. Although the source of the inefficiency here is somewhat different from the classic adverse selection models, the efficiency condition that is violated is exactly the same: Efficiency requires that consumers obtain insurance if and only if their willingness-to-pay for insurance exceeds their costs to insure.

Of course, it is an empirical question whether in employer plans it is possible to identify groups with different insurance demand, conditional on risk. I now turn to the data.

1.3 Data Source and Measurement of Risk

1.3.1 Data: Administrative Health Claims

Data on health plan selection and medical risk come from the administrative health claims of a single large employer.¹¹ As is typical among large, self-insured companies, the employer offered a menu of health plan options to its employees. Over the period studied, employees could choose between two PPO (Preferred Provider Organization) plans only.¹² Around 40,000 employees enrolled in these plans annually, and the claims data span 2004-2007.

Employees are traceable over time, and their plan elections each year are known. All contacts with medical service providers and all prescription drug purchases are observed. The claims files are line-item billing records. Besides prices and quantities, they list the portions of each invoice paid by the enrollee and by the health plan. Most records include one or more diagnosis or procedure codes, providing rich clinical information on health status. This clinical information is used to construct a measure of expected health risk that corresponds to r in the theoretical discussion of the last section.

Employees represent only 39% of all health plan enrollees, with spouses and children comprising the rest. 31% of employees enroll no dependents. These employees may nonetheless have spouses and children who receive insurance through other channels, such as the spouse's employer. The main estimation sample will be limited to the roughly one third of employees who enroll only themselves, since dependent enrollment complicates the analysis of demand differences by age and gender, as families are mixed demographic groups but must make a common health plan choice. Health plan data exist for enrollees up to age 64, though to avoid the complication of enrollees aging into Medicare and out of the sample over the panel, I cap the sample at age 59.

Table 1.1 displays worker characteristics. The first column includes all employees

¹¹These claims files were collected and purged of identifying information by Medstat.

¹²According to the Kasier Family Foundation Survey of Employer Health Plans (2007), employers are most likely to offer their workers a PPO or POS plan.

who are enrolled in a health plan. The second column is limited to employees who enroll no dependents. Workers are predominately male. Most are hourly workers. The row labeled “enrolled t-1” indicates the persistence of enrollment from year to year. It gives the fraction of current-year enrollees who were also enrolled in the previous year in *any* health plan offered by the company. Employment records, as distinct from enrollment records, are unavailable, but it is likely that turnover in health plan enrollment is largely driven by turnover in employment. Although no information is known regarding employees who waive coverage, data from the Kaiser Family Foundation (2007) Survey of Employer Health Benefits indicates that 15% of workers in firms of similar size and in the same industry waived coverage around this time. The firm is a manufacturer of consumer non-durables, with plant locations throughout the US.

Employees were offered two PPO options that differed in their cost-sharing rules—that is, the degree of consumption smoothing across health spending shocks they provided. Call these options L and H for low and high coverage. Contract H had a lower deductible, lower coinsurance, and a lower out-of-pocket maximum. Plan characteristics are listed in Table 1.2. The coinsurance rate in Table 1.2 is the marginal price faced by the consumer once the deductible has been met. It is 10% in plan H and 20% in plan L. Both the deductible and coinsurance are counted towards the out-of-pocket maximum. The cost sharing rules in Table 1.2 apply only to in-network services. Out-of-network services involve higher cost sharing in both plans, but these types of claims comprise a small minority of all claims. Both plans have an allowance for well-visits at no out-of-pocket cost to the enrollee. Prescription drug benefits are the same between plans in every year, except for a temporary change to drug copays in 2006.¹³

The employee contributions to plan premiums are not available in the data, so the

¹³No administrative data is available on the physician networks associated with each plan, but the networks appear to be similar or the same: Comparison of physician IDs in the claims data reveal that most (70%) of the primary care physicians visited by at least one Plan H enrollee were also visited by at least one Plan L enrollee.

prices of these plans are not known. Average plan costs, in contrast, can be directly and precisely calculated by aggregating the total health plan payouts. When prices are needed in the choice model below, they are constructed by applying typical employee contribution formulas to average plan costs. Fortunately, identifying demand differences across groups hinges only on observing the plan choices of employees who face the same prices, regardless of what those prices are.

Table 1.3 provides a high-level summary of employees' plan choices and healthcare consumption. The first column tallies the fraction of enrollees who choose plan H. Healthcare consumption in columns (2) through (4) is measured as the total bill paid to service providers. Column (2) gives the average overall expenditure. Columns (3) and (4) are conditional on the plan chosen. The first row lists these aggregate statistics for the entire estimation sample, pooling years 2005-2007. It is immediately apparent that adverse selection is reflected in the plan choices: Overall, 32% of enrollees choose plan H, and these enrollees consume twice as much healthcare on average as plan L enrollees. Moving down the table, the statistics are repeated for different demographic sub-groups. The same general pattern of adverse selection exists within each group as it does in the sample overall, but the level of costs vary widely across the groups. For instance, the average consumption of a 50-59 year-old is around 3.5 times that of an 18-29 year-old (\$6184 versus \$1702). *Prima facie*, this implies large subsidies from young to old, since premium contributions are regulated to be identical for all similarly-situated employees.¹⁴ Further, the young cannot escape these transfers by choosing the lower-price plan. Even conditional on choosing contract L, utilization in the oldest group is still 3.5 times that of the youngest group. Men also tend to subsidize women for the same reason: men's health spending is around \$1500 lower on average than women's spending in the sample.

What is particularly striking about columns (3) and (4) in Table 1.3 is how dramatically

¹⁴It is possible that younger employees are at least partially compensated by the employer for this apparent transfer, perhaps in other features of compensation or work environment.

average costs in *both* plans increase with age. These summary statistics are not easily reconciled with the assumption that, on average, young and old employees with the same expected medical costs make the same plan choices. For example, it would not appear to be the case that all enrollees share some common cutoff, choosing H if their expected costs are above that cutoff level, and choosing L if their expected costs are below it. To see this, note that average healthcare utilization among 18-29 year olds in the higher-premium plan (the plan that is cost-minimizing for those with the highest health claims) is only around \$3000, while the average utilization among 50-59 year olds in the lower-premium plan (cost-minimizing for those with the lowest health claims) is around \$5000. A common cutoff rule would tend to produce expected costs in plan H that are higher than expected costs in plan L, within and across subgroups. But columns (3) and (4) show the opposite when comparing young and old.

Table 1.3 shows that older enrollees are less likely to select plan L, but conditional on selecting L, their costs are relatively high. These patterns suggest that while the probability of selecting plan H increases in age, this is not simply due to the fact that risk increases in age. Instead, it appears that older enrollees, as a group, are more likely to choose plan H at all levels of risk. The summary statistics are merely suggestive on this point. More systematic analysis would compare plan choices between young and old enrollees who face the same *ex ante* level of expected healthcare costs. Doing so requires a measurement of expected health costs that can be used for the comparison. I now turn to constructing such a measure.

1.3.2 Risk Measurement

In studies of plan choice in health insurance, risk is most often proxied by age, gender, and limited or self-reported information on health status. Here, the claims records include every contact with a physician or hospital and all prescription drug purchases, enabling a more precise measure of risk to be constructed.

I use the Johns Hopkins Adjusted Clinical Grouper (ACG) to aggregate this rich cost and clinical information into a risk index.¹⁵ The ACG tool is widely used by insurers and managed care organizations to predict future spending, to flag high cost patients for intensive management, and to monitor drug adherence. Here, it is used solely to predict expected medical consumption, in dollars, on the basis of past episodes of care. The prediction is for total healthcare consumption—that is, the total payments made on behalf of the enrollee.

The ACG methodology can capture trajectories in healthcare utilization that would be missed by simpler methods. For instance, it treats a large expenditure on an injury differently from the same size expenditure on chronic disease management. Because the ACG prediction is based on the clinical information available at the time that employees make enrollment choices for the following year, it corresponds well with the information set assumed in the choice model below.

Only a few recent studies, including Carlin and Town (2008); Bundorf et al. (2008); Handel (2010), have been able to use this type of detailed risk control, since the data requirements are stringent. Precise controls for risk are important here, in part because a novel finding of this study is that there are large differences in insurance demand across age, conditional on risk. This pattern has necessarily been missed in studies that proxy for risk with age and gender.

The data permit a direct test of the ACG prediction by comparing it to the *ex-post* realized claims in the plan period. In the next section, I show that the ACG predictor accurately forecasts future expenditure: a regression of actual use on predicted use has a coefficient near one and an unconstrained intercept not different from zero. Alternative risk measures, including the predicted values from my own regressions of current expenditure on past expenditure (run separately by age), produce similar results to those shown in the

¹⁵Inputs into the ACG from my data include age, gender, pharmacy spending, medical (non-pharmacy) spending, inpatient stays, outpatient visits, ER visits, all ICD-10 diagnosis codes, all CPT procedure codes, national drug codes of all prescriptions filled, and places of service.

next section. However, less of the variance in realized healthcare costs is explained by these alternative predictors.

1.4 Reduced Form Results

1.4.1 Demand Patterns Across Age and Gender

In this section, I present evidence that willingness-to-pay for insurance varies strongly with age. For now, the intention is merely to document the broad patterns of demand while imposing minimal economic and econometric assumptions. Later, the welfare consequences of these demand patterns will be estimated in an expected utility framework that is better connected to theory, but also more reliant on structural assumptions.

Figures 1.3 and 1.4 plot local regressions of plan choice versus health risk, estimated separately by age. Health risk, predicted by the Johns Hopkins ACG tool, runs along the horizontal axis. The dependent variable is an indicator for choosing plan H. In Figure 1.3, health risk is measured as expected healthcare consumption in dollars, and the axis is scaled in logs. In Figure 1.4, this risk variable is adjusted to exclude anticipated drug expenditure. Both plans offer the same drug coverage, implying that drug purchases may not be an important consideration in plan choice. Both figures include data from years 2005 through 2007, and the sample is limited to the 1st through 99th percentiles of predicted cost within each age group.

The plots show that conditional on health risk, take-up of plan H increases fairly monotonically in age. By taking any vertical slice in Figures 1.3 and 1.4, one can compare the average choices of employees of different ages who face the same prices and carry the same expected costs. For instance, at \$1500 of expected healthcare consumption, which is near the median for 18-29 year-olds, the take up of plan H differs by 13 percentage points (54%) between the oldest and youngest groups, with 30-49 year-olds falling roughly in the middle. Another way to see these dramatic differences is to consider that while take

up rates reach 40% at \$2000 of predicted health expenditure for 50-59 year-olds, take up reaches 40% only at \$8000 of predicted expenditure among 18-29 year-olds.

Interpreting these results in terms of willingness-to-pay is straightforward: For any level of health risk, the take up rate is just the fraction of employees for whom willingness-to-pay for plan H over plan L exceeds the difference in employee contributions to premiums between plans L and H. The figures parallel the earlier theoretical analysis in Figure 1.2, which plotted willingness-to-pay against health risk. Since at all levels of health risk, a greater fraction of older employees choose plan H, the figures suggest that willingness-to-pay for plan H increases in age, after controlling for the cost of insuring the enrollee.

Conditioning on expected healthcare utilization most closely follows the model presented above, but these patterns are robust to conditioning on alternative measures of healthcare costs. Figure 1.6 replicates Figure 1.3, but uses last period’s medical expenditure along the horizontal axis. Figure 1.7 uses the *ex-post* realized spending in the period for which the plan choice is made. In both cases the patterns similar to the main results.

Figure 1.5 repeats the same graphical analysis by gender, using the the measure of predicted consumption that includes drugs. Women seem to have a higher willingness-to-pay for the more comprehensive insurance option than men, but demand differences across gender are far less pronounced than across age. Moreover, the apparent differences may be masking an age effect, since women are slightly older in the sample.

To better differentiate age and gender effects and control for other covariates, Table 1.4 reports results from a series of plan choice regressions that parallel the figures. The specification here is linear probability for ease of interpretation. Regressions include controls for all observable worker characteristics, including union status, whether full time or part time, whether paid hourly, and state of residence. These controls are potentially important—though in practice they don’t materially alter the results—because regulations require only that *similarly-situated* workers face the same prices and plan options, which means that employee contributions to premiums may differ by pay status, union status, or

plant location, for example. Since multiple years are included, standard errors are clustered at the person level, and year effects are also included in the control variables.

Consistent with the figures, the regressions in columns (1) and (2) confirm there is a large plan choice gradient in age, after conditioning on the additional controls. Coefficients on the age dummies in columns (1) and (2) can be interpreted as percentage point differences in take-up rates relative to the excluded category, 18 to 29 year-olds. Again, the take up patterns are monotonically increasing in age. Because the base take up rates are quite low, these correspond to large percent differences. Gender differences in willingness-to-pay (in column 2) are smaller, but indicate that women have a preference for more comprehensive insurance. Interactions of age and gender, not included in the table, were found to be insignificant. A novel finding of this paper is that willingness-to-pay for health insurance is very different between employees of different ages, and slightly different between men and women. (Follow-up work in other data sets and health insurance settings shows similar patterns. Among other firms whose data are available via Thomson Reuters, including the firm studied by Kowalski (2009), I find striking differences in plan choices by age conditional on measures of health risk.)

The analysis thus far has conditioned only on objective measures of health risk. It has ignored other variables that are presumably important in the consumer’s decision process, such as subjective risk assessment, risk aversion, or liquidity constraints. This is because these regressions are not intended to carefully represent the consumer’s choice problem. The point of view is rather one of a social planner or risk-neutral insurer, both of whom are concerned with plan choice as a function of cost to insure, but not necessarily with the preferences or budget constraints that generate demand. Einav et al. (2010a) summarize this point about sufficient statistics in welfare analysis in the following way: “Different underlying primitives (i.e. preferences and private information...) have the same welfare implications if they generate the same demand and cost curves.”

Here, age is merely a stand-in for a variety of unobserved parameters and decision

factors that influence willingness-to-pay, but as the model in Section 2 showed, it can be a relevant stand-in if it identifies groups with different demand curves, because it can then be used to more efficiently price discriminate.

1.4.2 Does the Econometrician Observe a Sufficient Statistic for Expected Costs?

The regressions in Table 1.4 are intended to determine whether age has residual power to predict plan choice once insurer's expected costs controlled for. Of course, age is not exogenous in the regression, as it is likely correlated with many unobserved variables that are incorporated into the error term of the regression. Nonetheless, this correlation does not threaten the validity of the claim that conditional on objective risk, the old are more likely to choose more complete insurance than the young, which is one of the main empirical claims of the paper. Validity *is* threatened, however, if bias in the health risk control creates bias in the age coefficients. In this section, I provide evidence on the quality of the ACG prediction of health risk.

One straightforward way in which results from the plan choice regressions could be misleading is if there were bias in the ACG prediction that differed systematically by age. For instance, the ACG could under-predict health risk for the young and over-predict it for the old. If so, it could produce spurious age effects in the regressions.

A more subtle problem would occur if the ACG prediction had no differential bias across groups, but suffered from measurement error that biased down the coefficient on risk in the plan choice regressions. In that case, estimates of the age dummies would be biased as well, in the sense of no longer representing age effects that are independent of expected costs. For example, in the extreme, if the health risk variable were purely noise, one would expect large age coefficients even if the true age effects were zero, because the age coefficients would simply capture the correlation between age and average risk.

Fortunately, the data allow a direct test for various types of error in the health risk

variable. Recall that the ACG predicts expenditure in the plan period based on expenditure in the previous period. For example, predicted expenditure for 2006 is based on utilization in 2005, including prescription drug ingredients, diagnosis and procedure codes, inpatient stays, and physician visits. Since actual expenditure in 2006 is observed, the fit of the ACG prediction can be assessed directly by regressing the actual expenditure in 2006 on the ACG prediction of expenditure for 2006. If the ACG constitutes a sufficient statistic for expected spending, that regression should yield a slope of one and an unconstrained intercept of zero. If the slope is less than one, we might expect a measurement error problem. If the slope and intercepts differed when the sample is restricted to different age groups, it would indicate a differential bias problem, in which the precision or accuracy of the prediction varied by age group. Either would be problematic when the ACG measure is used in the plan choice regressions.

Table 1.5 displays the results from a series of regressions of actual healthcare spending on ACG predicted spending. The regression in column (1) includes all enrollees in the main estimation sample. The estimates suggest that measurement error is not causing an attenuation bias problem: The coefficient on the predicted expenditure (the slope) is very close to one, and the intercept is close to zero. Attenuation bias due to measurement error, if present, would manifest as a slope less than one in this regression. Thus the ACG predictor gives an accurate measure of expected costs for enrollees as a whole.

Columns (2) through (6) address the question of whether the ACG is differentially biased or error-prone across the age groups. Column (2) adds age dummies, which are insignificant when the ACG measure is included in the regression. Columns (3) through (6) rerun the regression separately by age. In each of these four regressions, the point estimates of the slope generally remain within a standard error of one, and the intercepts lie close to zero. This is evidence against the concern that the observed age effects in the plan choice regressions are an artifact of differential quality in the health risk variable by age. Such a story would require that expected costs for older enrollees are grossly

underestimated. Here there is no large, positive intercept for the old, which would indicate an underestimate of risk. Nor is there a large, negative intercept for the young.

The regression in the last column in Table 1.5 includes an indicator for enrollment in plan H in the period for which the expenditure prediction is made. It is intended to jointly test for (1) a behavioral response to facing lower marginal costs of care (usually called moral hazard in this context), and (2) selection on private (to the econometrician) knowledge of health status. A positive coefficient would be consistent both with moral hazard (for a fixed health state, those in plan H consume more healthcare) and with selection on residual private information (those who know their health is worse than the econometrician observes are more likely to optimally choose plan H). The two phenomena are difficult to separately identify in observational data.¹⁶ Either would indicate that the relationship between predicted costs, realized costs, and underlying risk is more complex than so far assumed, though not necessarily in a way that upsets the finding of differential insurance demand by age. The positive and significant coefficient on the plan H indicator implies that people who choose plan H in year t cost \$900 more than predicted in year t .

It is important to recognize that if the coefficient on the plan H dummy is indeed due to moral hazard, it biases against finding the observed higher willingness-to-pay among the old: Because the marginal cost of consuming health care is lower under plan H, consumers in H may seek additional care. Since expected health costs are estimated on the basis of past costs, the health risk prediction is potentially higher for the same person if she is observed in plan H in $t - 1$ than if she is observed in plan L in $t - 1$. Further, because the old are more likely to take up H in $t - 1$, the ACG may differentially overestimate risk for the old. In that case, the true relationship between risk and plan choice for the old in

¹⁶Some headway might be made in disentangling moral hazard and residual private information by controlling for last year's plan choice and current year's plan choice simultaneously. Unfortunately, these are highly correlated since plan choices are persistent, and the difference between them is identified by switchers. Switchers are a tiny fraction of the sample and are exactly the group most likely to have residual private information about impending changes in healthcare needs, implying they are not particularly informative of the broader fit of the ACG prediction.

Figure 1.3 would have to shift left (to lower risk) to correct for this overestimation. But this would just widen the gap between the local regression lines of the young and old.

If the coefficient on the plan H dummy is due to residual private information, it could bias toward finding effects. However, there are reasons to believe this is not a first order concern. First, the magnitude of the potential private information effect is small relative to the observed demand differences. Recall that the while take-up rates for plan H reach 40% at \$2000 of predicted health expenditure for 50-59 year-olds, take up reaches 40% only at \$8000 of predicted expenditure among 18-29 year-olds. Second, residual private information is only a serious problem here insofar as it reflects *differential* bias in the expected costs variable between age groups. But columns (2) through (5) of Table 1.5 show there is little support for any hypothesis that claims differential bias in the ACG prediction by age.

The importance of the set of regressions in Table 1.5 is that they argue against a large set of stories in which the observed relationship between age and plan choice is driven by some artifact of the risk variable. Overall, Table 1.5 suggests that expected costs are appropriately controlled for by the ACG predictor, so that the age coefficients are not reflecting unobserved components of risk. A related issue, but one with less important consequences for the estimation exercise, is whether the agent knows her own risk. This is taken up in the next section.

1.4.3 What Drives Demand Differences?

Describing the origins of differences in plan choice patterns across demographic groups is not the goal of this paper, nor is it critical for the welfare analysis. Nonetheless, the differential patterns of health plan selection by age beg the question of what drives demand differences across demographic groups.

Differences in risk preferences, subjective risk assessment, and income are all plausible explanations. Unfortunately, there is very little scope to differentiate between these alternatives with the data at hand: There are no independent measures of risk aversion for this

set of employees. Nor are there any survey questions that might elicit employees' subjective assessments of health risk. Employee income is likewise unknown. I am, however, able to provide some suggestive evidence on whether default options and/or inertia in plan choice play a role in explaining the phenomenon. An important avenue for future work will be to more carefully establish the source of the demand differences highlighted in this paper. For now, I simply note the classes of plausible explanations, and the types of policy changes to which the observed demand patterns would be invariant.

The type of counterfactual policy considered below is one that allows employee premiums to differ by age, reflecting differences in average costs. In practice, this means lowering prices for younger workers and raising them for older workers. If demand differences were driven by underlying differences in deep parameters such as risk aversion, then the willingness-to-pay patterns would survive most plausible policy changes, including the one considered here. The demand differences between young and old are also likely to be invariant to the new pricing policy if they originate from differences in liquidity constraints or from differences in subjective assessments of risk. The former may arise because younger workers have fewer dollars to commit to health spending at the time of plan enrollment. The latter could arise if enrollees only pay attention to the portion of bill they pay, such as the deductible and coinsurance, and have little knowledge of the total bill paid on their behalf. If the old have more experience, their subjective probabilities of healthcare utilization may be more accurate.

Another class of hypotheses for explaining the demand patterns involves behavioral economics phenomena such as sensitivity to default options and inertia in plan choice. The young and old may be differentially active or passive in their annual plan choices, or they may have begun their tenure during a period when defaults were different. This set of explanations would suggest that demand differences by age are not very robust to interventions and changes in the policy environment.

Some insight into whether these types of behavioral phenomena play a role in the

observed relationship between age and insurance demand can be gained by restricting the analysis to a smaller sample of employees who are entering their first year of health plan enrollment. In the first year, switching costs are small or zero, by most definitions. It is unknown whether a default option existed for the first year of health plan enrollment. Regardless, it is reasonable to believe that initial enrollment is a time during which typical employees are making their most active decisions.

Expected healthcare consumption cannot be calculated for first year enrollees, since they have no medical history to use as a basis for prediction. Therefore, in columns (4) and (5) of Table 1.4, I repeat the earlier plan choice analysis, but limit the sample to only first-time enrollees, using *ex-post* realized consumption from their first year as the health risk variable. Unlike predicted healthcare expenditure, actual expenditure includes some zeros, so the expenditure variable is not immediately amenable to a log transformation. I take two ad-hoc approaches to the transformation. In column (4), I drop the zeros. In column (5), I add \$50 to total claims before taking logs. The difference in take-up of plan H between the 18-29 year-olds and 50-59 year-olds in both specifications is very similar to in the main results, though the pattern is slightly different in the middle ages. These results are suggestive that the demand heterogeneity across age is present even at initial enrollment. The differential enrollment patterns are therefore probably not driven by older employees “aging in place,” by older employees entering employment at a time of different default options, or by older employees having more experience with the employer’s health plans.

1.5 Structural Estimates

1.5.1 Discrete Choice Model

The reduced form results showed large differences in insurance demand across demographic groups. The model developed in Section 2 suggests that these patterns, when combined

with restrictions on price discrimination, can lead to inefficient insurance allocations. I now turn to a structural discrete choice model that allows statements about redistribution and welfare losses to be made more precise. This precision comes at the cost of imposing the assumption of an expected utility decision process.

I estimate a model of health plan selection that follows Einav et al. (2010b) and Handel (2010). The choice between health plans is modeled as a choice between money lotteries, in which the fuller coverage plan (H) provides more smoothing of health expenditure shocks at a higher up-front price. Employees choose the plan that maximizes expected utility.

As is standard in the insurance literature, I assume CARA preferences. The von Neumann-Morgenstern CARA utility index takes the form

$$u(x) = -\frac{1}{\gamma}e^{-\gamma \cdot x}, \quad (1.2)$$

where γ is the risk aversion parameter, and x is used as a placeholder for the money lottery term. To apply the utility index to this setting, I replace x with the payments that agent i faces in each state of the world, conditional on choosing plan j . This term consists of initial wealth, W ; the employee's contribution to the plan premium, P ; and the out-of-pocket expenditures spent satisfying deductibles and coinsurance, OOP :

$$u(W_i - P_j - OOP) = -\frac{1}{\gamma}e^{-\gamma \cdot (W_i - P_j - OOP)}. \quad (1.3)$$

The basic form of the utility index to be estimated is presented in eq. (1.4). Expected utility integrates this index over states of the world defined by out-of-pocket medical costs, which are a function of uncertain medical need. The expected utility of plan j for a person facing the risk distribution F_{ij} is

$$E[V_{ij}] = \int -e^{\gamma(P_j + OOP - \delta_o H - \sum_a \delta_a age \times H - \delta_F Fem \times H + \epsilon_{ij})} dF_{ij}(OOP). \quad (1.4)$$

The subscript i indexes individuals, and $j \in L, H$ indexes plans. Agents choose the plan that yields the highest expected utility, given their preferences and the risk they face. Moving from eqs. (1.3) to (1.4), I've dropped the $1/\gamma$ term in front, as well as $e^{\gamma W_i}$ (which can be factored out), because these terms are constant over all states of the world and plan choices, affecting only the scale of utility. For the same reason, only the price difference between plans L and H is needed for estimation, as the common component of price can be factored out. H is an indicator for plan H, and the vector $age \times H$ is a set of interactions between plan H and age group. $Fem \times H$ is an interaction between female and plan H. The coefficients on these interactions are the main estimates of interest, analogous to the age and gender dummies in the Table 1.4 regressions. They are intended to capture preferences for plan H that differ systematically across demographic groups. Because they enter the utility function in the money term, they measure plan preference differences in dollars.

Unobserved decision factors are represented by ϵ_{ij} , which is a plan-specific, normally distributed error term with mean zero:

$$\epsilon_{ij} \sim N(0, \sigma_{e_j}^2).$$

The variance of the error, $\sigma_{e_j}^2$, is normalized to zero for plan L, since only the difference in error terms between the two plans is identified.

Besides the analytic convenience of using a CARA utility index, it is a necessary simplification in this setting since wealth is not observed. An additional benefit of assuming CARA is that because it is a typical assumption in the insurance literature, it allows direct

comparison of my estimated risk aversion parameter with other recent work.

In eq. (1.4), the distribution $F_{ij}(OOP)$ has an individual- and plan-specific subscript because it reflects both the idiosyncratic health risk faced by the employee and the plan-specific cost sharing rules that translate healthcare consumption into out-of-pocket expenditure. It is estimated separately and prior to the estimation of the parameter vector, as described in the next subsection.

The full parameter vector to be estimated is $\theta = [\gamma, \delta_o, \delta_a, \delta_F, \sigma_{eH}]$. In many settings, the data requirements for estimating this type of expected utility model would make the exercise infeasible. Insurance plan choice is a recently exploited exception. Insurers' administrative data allow the careful measurement of risk. Moreover, the choice problem consumers face in selecting insurance is at least potentially well-described by a choice over money lotteries. This is particularly true in cases such as this one because only net payments—and not provider access or other plan features—differ between the insurance options.

1.5.2 Expected Costs

As an input to the choice model, I estimate two out-of-pocket risk distributions for each employee, which correspond to the money lotteries implied by the two insurance plans. These distributions, F_{ij} , are constructed in two steps. First, I estimate a person-specific distribution of medical utilization risk, G_i . This is a distribution over the total dollar amount paid to providers on behalf of the enrollee in the plan year. Second, plan-specific mapping functions translate the raw utilization distributions G_i into out-of-pocket costs distributions F_{ij} .

Estimating G_i begins by creating cells defined by age and small bands of predicted use. Predicted use is generated by the Johns Hopkins ACG software, as described above. Enrollees in the same cell are assumed to face the same distribution of healthcare risk. Under this assumption, the *ex-post* distribution of healthcare consumption within a cell

identifies the *ex-ante* distribution of risk faced by all enrollees within that cell. For each cell, I separately parameterize the distribution of realized healthcare consumption via maximum likelihood. Risk is assumed to follow a mixture distribution consisting of a lognormal and a point mass at zero.¹⁷ Once G_i is estimated, deterministic mapping functions $H_j : G_i \rightarrow F_{ij}$ convert healthcare utilization into out-of-pocket costs according to plan rules.

When selecting cell bands, there is a tradeoff between keeping the cells as tightly defined as possible and having a sufficient within-cell sample size to precisely estimate the parameters. In the specification presented, there are 37 cells total. The minimum and median cell sizes are 129 and 211, respectively. Experiments in changing the size of cells yielded little impact on model estimates.¹⁸

Finally, a maintained assumption here is that healthcare consumption will be identical under either plan. This is equivalent to an assumption of no moral hazard in the sense of price sensitivity.

1.5.3 Identification

In the data, there are no observed innovations in plan prices that could be used to trace the insurance demand curves in a typical reduced form approach. This is the motivation for the expected utility model. The assumption of an expected utility decision process circumvents this difficulty because by framing the choice between two health plans as the choice between two money lotteries, it imposes parity between dollars spent on premiums and dollars spent satisfying coinsurance and deductibles. This expected utility assumption

¹⁷Lognormal with a point mass at zero is a commonly-used parameterization of medical risk. A mixture of a point mass with a two parameter Weibull—an alternative which is less widely used—was tried as well but tended to underestimate the right tail of the distribution.

¹⁸Alternative approaches to estimating F_{ij} include: (1) using the empirical CDF within each cell as G_i rather than fitting a mixture of standard distributions to the data, and (2) abandoning the cell-based approach and assuming all enrollees face a distribution of risk in which the distribution’s parameters are non-parametric functions of age and predicted use. I plan to explore these options in future versions of the paper, but since the results are insensitive to changes in the cell divisions, and since the estimated patterns of plan preference generated by the chosen parameterization are observed even in the reduced form results, it seems unlikely that doing so will result in significant changes in the paper’s findings.

exploits the substantial variation across enrollees in the “total price” of a health plan, which includes not only plan premiums, but also the out-of-pocket costs paid in deductibles and coinsurance, which are a function of idiosyncratic health. This is a strong assumption, and one that would not be robust to the type of behavioral economics story in which consumers view spending \$50 on a monthly premium differently from spending (a sure) \$50 on a hospital bill copayment.

In contrast, the identification of demand differences across age and gender does not rely on the assumption of expected utility. Instead, identification of demand differences is generated by the institutional feature that all similarly-situated employees face the same prices for the same insurance contracts. Because all agents face the same prices, comparisons of enrollment choices between young and old or male and female employees who face the same objective risk reveal demand differences across groups. Therefore, even though employees are observed facing only a single set of prices in the data, differential willingness-to-pay across demographic groups is cleanly identified at the particular prices they face.

The interactions of plan H with age and gender pick up effects that are not well-described or rationalized by an expected utility framework that assumes identical preferences across these groups. To the extent that young and old make systematically different choices, the residual power of age to predict plan choice loads onto these interaction terms.

An alternative specification would have allowed the risk aversion coefficient to vary by age, rather than including interactions of age and plan H dummies in the money term. There is some isomorphism between allowing the risk aversion coefficient to move with age and allowing the monetized preference for plan H to move with age—increasing either coefficient can rationalize a higher willingness-to-pay for plan H in the data. The effects are not perfectly collinear, however, because whereas the age dummies capture a preference for H that is uniform for all people in the age group, the risk parameter must also fit within-group differences in higher moments of the risk distributions $F_{ij}(OOP)$.

I use age dummies in the money term, rather than in the risk aversion term, for

two reasons. First, the dummies yield easily interpretable differences in valuation that are measured in dollars. Second, there is no corroborating evidence suggesting that willingness-to-pay differentials are caused by differential risk preferences, rather than by differential liquidity constraints, differential subjective probabilities, or anything else. Therefore, I wish to avoid confusion that might result by casting results in terms of risk aversion, a concept which is endowed with meaning outside of this estimation exercise.

Finally, plan costs are observed, but employer contributions to plan costs are not, which means that the price difference between contracts L and H that employees face is assumed, not known. In estimation, price enters identically to the plan H intercept, so changing the assumed price difference changes the estimate of the plan H intercept dollar-for-dollar, but no other coefficients. The intuition behind this isomorphism is that the plan H term effectively measures how, conditional on a risk aversion parameter and their objective risk, agents under- or overvalue contract H relative to its price. As the assumed price changes, so must this term. I discuss the implications of this in detail in the results section.

1.5.4 Estimation Details

Simulated Maximum Likelihood Estimation (SMLE) is used to estimate eq. (1.4). Given the choice model outlined above, there is no closed-form expression for the choice probabilities. Therefore simulation is needed to calculate likelihood values. Here, I assume a normal error that enters the model non-linearly, but even if the error structure admitted a closed form solution to the choice probabilities (as would be the case, for example, with an additive extreme value error, which yields logit probabilities), simulation would still be necessary in order to numerically integrate over the risk distribution $F_{ij}(OOP)$.

Estimation begins by fixing the parameter vector θ and taking Q draws from the distribution of out-of-pocket expenses, $F_{ij}(OOP)$, to calculate an expectation over health risk. Conditional on a set of values from the parameter vector (including a single draw of

the error term ϵ_{ij}), the expected utility for person i in plan j is:

$$E[V_{ij}^r] \approx \frac{1}{Q} \sum_{q=1}^Q -e^{\gamma \cdot (P_j + OOP^q - \delta_o H - \sum_a \delta_a age \times H - \delta_F Fem \times H + \epsilon_{ij}^r)} \quad (1.5)$$

The superscript r denotes a draw from the normal error term.

A basic accept-reject (AR) simulator would assign a value of zero or one as the choice probability of plan j depending on whether j generated the highest utility given the simulated value of the error, ϵ_{ij}^r . Here, a smoothed AR simulator is used instead. For each draw r , a smoothing function is applied to ensure that the probability falls strictly between zero and one, avoiding problems otherwise arising from flat portions of the log likelihood function. I employ a logit smoothing kernel that follows Train (2009):

$$S_{ij^*}^r \equiv Pr(j = j^* | \theta^r, n) = \frac{\exp(E[V_{ij^*}^r]/\lambda)}{\sum_j \exp(E[V_{ij}^r]/\lambda)}. \quad (1.6)$$

This smoother approaches an unsmoothed AR simulator as λ approaches zero.¹⁹ The output of the smoother is the probability of selecting a plan, conditional on values from the parameter vector, including a single draw from the normal error term. In order to calculate the simulated choice probability, many draws are taken from the error (still for a fixed parameter value of σ_ϵ). Averaging the smoother values across these draws yields the simulated probability of person i choosing plan j :

$$\check{P}_{ij}(\theta) \equiv \frac{1}{R} \sum_{r=1}^R S_{ij}^r(\theta). \quad (1.7)$$

¹⁹The appearance of logit formula at this step has nothing to do with assumptions on the error term—it is purely a convenient way to ensure that choice probabilities fall strictly between zero and one.

Finally, the simulated log-likelihood function is defined by

$$SLL(\theta) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J d_{ij} \ln \check{P}_{ij}(\theta), \quad (1.8)$$

where d_{ij} is an indicator function for the observed choice of each individual. Standard numerical maximization routines are used to search for the maximizing parameter vector. In practice, I set $Q = 50$, $R = 100$, and $\lambda > 5$.

1.5.5 Parameter Estimates

Table 1.6 lists estimated parameter values for the choice model. Consistent with the patterns of selection found in the reduced-form analysis, willingness-to-pay for plan H increases monotonically in age. The coefficients indicate that, after controlling for objective risk, 30-39 year-olds value the more comprehensive PPO option by \$261 more than 18-29 year-olds. 40-49 year-olds value it by \$323 more. 50-59 year-olds value it by \$508 more. Willingness-to-pay is also somewhat higher among women.

The plan H intercept indicates that overall, agents make choices as if contract H were priced at \$955 more than the assumed price. There are several ways to interpret this coefficient. First and most obvious, actual employee contributions to premiums are not observed, so the assumed price difference between contracts L and H may be wrong, and the coefficient may partially reflect this. Alternatively, the CARA model may fail to adequately describe all the complexities of health plan choice, in which case, the plan H coefficient is merely a convenient modification that allows the model to better fit choice patterns in the data. In terms of fit, a likelihood ratio test strongly rejects the null model in which the plan H intercept is not included ($p < .0001$).

The estimated risk aversion coefficient is $6.96 \cdot 10^{-4}$. Since this number is not readily interpretable, in Table 1.7, I compare it to estimates of risk aversion from other studies.

Risk preferences are likely to be context-dependent, and several of the estimates listed in Table 1.7 are derived from other insurance markets. It is nonetheless of interest that my estimate falls squarely within the range of the existing literature. My estimate comes closest to estimates from Handel (2010), which were derived in a similar context (employee health plan choices) though with a different identification strategy. To further aid interpretation of the risk aversion coefficient, the last column in Table 1.7 displays a certainty equivalent measure of the risk parameters. The certainty equivalent here is the amount X that would make someone just indifferent between accepting a gamble in which they win \$100 or lose X with equal probability versus a status quo where nothing happens. The value X implied by my estimate of risk aversion is around \$93.

1.6 Policy Simulations

1.6.1 Welfare Metric

Quantifying welfare changes under counterfactual pricing policies requires a welfare metric. I follow Einav et al. (2010a) in calculating surplus via a certainty equivalent. For each individual, the certainty equivalent, CE_{ij} , is the money amount that makes agent i just indifferent between paying that sure amount and facing the uncertain loss associated with insurance plan j .²⁰ Though health insurance smooths consumption against health risk, the insurance contracts still involve some uncertainty in terms of out-of-pocket payments. Total surplus is defined as the difference between the certainty equivalent associated with the *chosen* insurance plan j^* and the certainty equivalent of having no insurance, minus the insurer's cost of enrolling the individual in the chosen plan. Let the superscript *cf*

²⁰ CE_{ij} is found by equating the utility of the sure amount to the expected utility of the lottery, and solving for the sure amount. Recall that expected utility is specified as a CARA utility index integrated over the distribution of health expenditure risk. Thus the certainty equivalent is the value CE_{ij} that satisfies: $-\frac{1}{\gamma}e^{-\gamma \cdot CE_{ij}} = \int -\frac{1}{\gamma}e^{\gamma \cdot (P_j + OOP - \delta_o H - \sum_a \delta_a age \times H - \delta_F Fem \times H + \epsilon_{ij})} dF_{ij}(OOP)$. This equation is solved for CE_{ij} and calculated for each individual by applying the estimated parameter values and the individual-specific distribution of health expenditure risk, $dF_{ij}(OOP)$.

denote the counterfactual, and let j^{*cf} indicate the plan chosen under the counterfactual price. Denoting individual total surplus (welfare) as W_i , the change in total surplus under counterfactual pricing rules is:

$$\Delta W_i = \underbrace{(CE_{ij^{*cf}} - Cost_{ij^{*cf}})}_{\text{new surplus}} - \underbrace{(CE_{ij^*} - Cost_{ij^*})}_{\text{old surplus}} \quad (1.9)$$

Since the certainty equivalent measures willingness-to-pay for insurance, it is straightforward to interpret the welfare metric in Eq. (1.9) in light of the efficiency definition in Section 2. There, efficiency required choosing the plan for which willingness-to-pay was in greatest excess of the insurer's cost of providing insurance. Here, welfare improves under counterfactual pricing for an individual if and only if his gain in utility under the newly-selected plan is larger than the increase in the insurer's cost to cover him in the new plan. Note that the portions of the premium price paid by the employee and by the firm do not enter here, with the implication that net welfare is not altered simply because an employee pays more or less for a plan.

1.6.2 Counterfactual Pricing

Three counterfactual pricing policies are considered here. Given the large subsidy applied to plan L under all pricing schemes considered, all enrollees are assumed to choose at least plan L. Thus, the relevant price in this exercise is the difference in employee contributions between plans L and H. In all counterfactuals, the welfare gains from new prices are compared to a baseline in which the employer sets employee premiums equal to 20% of the average plan costs (an 80% subsidy rule). This level of subsidization is typical for a large employer (Kaiser Family Foundation 2007). Only prices, and not the menu of options, are assumed to change.²¹

²¹A stable menu of plans is a common assumption in the literature that estimates welfare effects from alternative pricing of health insurance contracts.

The first counterfactual pricing policy considered is optimal marginal cost pricing as described in Cutler and Reber (1998) and elaborated in Carlin and Town (2008). This marginal cost pricing rule sets the single uniform price difference between plans L and H equal to the difference in costs of enrolling a marginal employee in H rather than L. Marginal here means an employee who is just indifferent between choosing either plan at the given price, regardless of her expected costs to insure. This pricing rule yields constrained efficient prices, subject to a uniform pricing restriction. This counterfactual has been reported in other papers, allowing a comparison to other recent studies.

The second pricing policy considered is a variant on the Cutler and Reber (1998) marginal cost rule, applied separately by age group in the same four age bins used throughout the analysis. Comparing this counterfactual to the first gives some insight into the additional gains of allowing age-adjusted pricing even after perfectly correcting for the inefficiently high uniform price created by adverse selection.

The third counterfactual pricing policy is the simple and common employer subsidy rule of covering an equal percentage of all plan costs, but applied separately by age group. I use 80% as the subsidization rate, just as in the baseline. This implies that employees who select H pay 20% of the difference in average costs between plans L and H, where those average costs are calculated over people in their own age group. Unlike the first two counterfactuals, it does not require that the employer or its insurance administrator observe the demand and marginal cost curves for its health plans. Since the information requirements are stringent for the marginal pricing rule (whether applied overall or within age groups), this provides a more feasible pricing alternative.

The counterfactuals are summarized below. The notation ΔP makes explicit that only the price difference between plans plan H and plan L matters here:

1. Marginal cost *uniform* pricing, following Cutler and Reber (1998):

$$\Delta P = \sum_{i \in N} \left(\frac{\partial Pr_i(H)}{\partial \Delta P} (\Delta Cost_i) \right) / \sum_{i \in N} \frac{\partial Pr_i(H)}{\partial \Delta P}$$

2. Marginal cost *age-specific* pricing, applying Cutler and Reber (1998) separately to each age group:

$$\Delta P_{agegroup} = \sum_{i \in agegroup} \left(\frac{\partial Pr_i(H)}{\partial \Delta P} (\Delta Cost_i) \right) / \sum_{i \in agegroup} \frac{\partial Pr_i(H)}{\partial \Delta P}$$

3. Average cost *age-specific* pricing, with an 80% subsidy to the average costs of each plan within each age group:

$$\Delta P_{agegroup} = .8(AC_H - AC_L)|_{agegroup}$$

1.6.3 Welfare Analysis

To simulate choices under the new prices, I replace the prices in the choice model and use the estimated coefficients and person-specific risk to find the utility maximizing option for each enrollee.

Table 1.8 displays pricing, enrollment, and welfare changes resulting from the three alternative pricing schemes. The first two columns report how the relative price of plan H changes under the counterfactual. Columns (3) and (4) report the old and new enrollment levels in plan H. Column (5) reports changes to welfare per enrollee per year in dollars. Column (6) scales these changes by average total healthcare expenditure in the sample.

Under all three counterfactual pricing rules, differences in premiums between plans L and H decline for enrollees of all ages, and more enrollees of all ages choose plan H, compared to the baseline. This implies that even with an 80% subsidy to plan costs, premiums are inefficiently high at baseline. (Note that smaller price differentials between plans L and H do not imply lower prices or larger subsidies overall. The prices of both plans can rise while the differential between them shrinks.)

Panel A in Table 1.8 shows that welfare could be improved by an average of \$11 per enrollee per year compared to the baseline by setting the uniform prices according to the marginal cost rule. This \$11 measures the adverse selection loss from sub-optimal uniform pricing compared to optimal uniform pricing. This same counterfactual was performed in

Carlin and Town (2008), Bundorf et al. (2008), and Einav et al. (2010a). In these papers the estimated welfare gains from optimal uniform pricing are \$13, \$17 to \$80, and \$6 per enrollee per year, respectively.

Panel B further refines the optimal pricing rule by applying it separately within age groups. The resulting optimal age-specific prices differ significantly from each other and from the optimal uniform price. For instance, the optimal price for the youngest group is \$285, while the optimal price for the oldest group is \$525. However, the additional welfare gain in Panel B compared to Panel A is small, just \$1 extra per enrollee per year. Despite significant differences in demand (and in the implied optimal prices) across groups, there appears to be very little welfare loss to pricing uniformly here.

Panel C asks how much of the potential gains of optimal *uniform* pricing can be captured by a simple, mechanical pricing rule in applied separately by age that requires knowing nothing about demand and cost curves. In Panel C, I simply apply the same baseline pricing strategy—an 80% plan cost subsidy—to each age group separately, rather than the whole enrollee population. It is interesting that almost all the gains (95%) of optimal Cutler and Reber (1998) pricing can be made up by this naive age-adjusted pricing strategy in this setting. The intuition for why this is the case is that risk pools here are less selected within age than overall. This was apparent even in the summary statistics of Table 1.3. Since adverse selection tends to distort choices in relation to the size of average cost differences between the risk pools, the age-specific prices are naturally less susceptible to distortion because the age-specific risk pools differ less across the plans.

1.7 Conclusion

This paper complements the recent empirical findings of substantial preference heterogeneity in insurance markets, first by documenting heterogeneity in demand for employer health plans and second by exploring the welfare consequences of prohibiting price discrimination when such heterogeneity exists. When preference heterogeneity is correlated with observ-

able characteristics, prohibiting price discrimination rules out potentially more efficient pricing schemes, and the classic solution to the adverse selection problem (subsidizing the adversely selected plan) no longer generates first-best allocations.

In the employer health plan setting I examine, I find that willingness-to-pay for more comprehensive insurance is significantly higher for older employees than for younger employees who face the same expected medical spending. Given these choice patterns, price discrimination on age could induce employees to self-sort themselves more efficiently than would a uniform price. Nonetheless, the welfare impacts of failing to optimally price on age are small, reinforcing the insight from the recent literature that suboptimal pricing in health insurance need not lead to meaningful welfare losses.

Despite small losses here, the size of welfare losses in other insurance markets and settings constrained by pricing regulations remains an empirical question. Many insurance markets are facing new community rating or anti-price discrimination regulations. For instance, the high court in the European Union ruled in 2011 that prices of car insurance and life insurance cannot differ by gender. In the US, the state health insurance exchanges established by the Affordable Care Act will prohibit price discrimination by gender and most other observables, but will allow (bounded) discrimination on the basis of age and smoking status. The general insight from this paper is that giving up the ability to price on observable characteristics can impact welfare not only because it removes the ability to price on correlates of risk, but also because it removes the ability to price on correlates of valuation. Therefore, there may be social objectives that make uniform pricing desirable, but prohibiting price discrimination comes at the cost of creating potential inefficiencies beyond the well-known adverse selection effects, and these inefficiencies cannot be corrected by any mechanism that sets uniform prices.

	All Employees	No Dependents
	avg 2005-2007	avg 2005-2007
employees	~34,749	~10,149
enrolled t-1	0.81	0.73
female	0.26	0.40
age		
18-29	0.14	0.26
30-39	0.30	0.25
40-49	0.35	0.28
50-59	0.20	0.21
hourly	0.68	0.67
union	0.12	0.14
full-time	0.96	0.94
region		
Northeast	0.14	0.17
North Central	0.26	0.28
South	0.49	0.44
West	0.11	0.11

Table 1.1: Characteristics of enrolled employees. The first column shows characteristics for the entire workforce, conditional on health plan enrollment. The second column is limited to employees who enroll no dependents, which comprise the main estimation sample. Except for the first row which gives counts, all numbers represent fractions. “enrolled t-1” counts the fraction of employees who were enrolled in any health plan one year ago. Regions follow census definitions.

	Plan years 2005 - 2007	
	PPO _L	PPO _H
Deductible	500	300
Coinsurance	20%	10%
OOP maximum	\$4250	\$2300
Copay ER	\$0	\$0
Free well visits	yes	yes
Copay drug	10/5	10/5
Coinsurance drug	30%	30%

Table 1.2: PPO cost-sharing rules for individuals, 2005-2007. The higher coverage plan, H, involves less cost-sharing in terms of deductible, coinsurance, and out-of-pocket maximum. The coinsurance rate is the marginal price faced by the consumer once the deductible has been met. Both the deductible and coinsurance are counted towards the out-of-pocket maximum. These cost sharing rules apply only to in-network services. Out of network services involve higher cost sharing in both plans, but these types of claims comprise a small minority of all claims. Both plans have an allowance for well-visits at no out-of-pocket cost to the enrollee.

	Fraction	Average Healthcare Expenditure (AC)			
	Choosing Plan H	All enrollees	in Plan L	in Plan H	obs
	(1)	(2)	(3)	(4)	(5)
overall	32%	\$3,612	\$2,555	\$5,834	30,447
by age					
18-29	21%	\$1,702	\$1,379	\$2,951	7,927
30-39	30%	\$2,759	\$2,077	\$4,350	7,756
40-49	36%	\$4,254	\$2,782	\$6,833	8,412
50-59	44%	\$6,184	\$5,026	\$7,655	6,352
by gender					
men	29%	\$2,874	\$2,069	\$4,840	18,235
women	37%	\$4,714	\$3,372	\$6,998	12,212

Table 1.3: Health plan selection and healthcare consumption, 2005-2007. Column 1 lists fraction of enrollees who choose plan H. Healthcare consumption is listed in columns 2 through 4 and is measured as the total bill paid to service providers. Column 2 gives the average overall expenditure. Columns 3 and 4 are conditional on the plan chosen. The first row lists these aggregate statistics for the entire estimation sample. In subsequent rows, the statistics are repeated for different demographic sub-groups. The levels of healthcare expenditure vary widely across age groups, which implies large subsidies from young to old. Adverse selection is present in the plan choices: Overall and within age groups, enrollees choosing plan H consume significantly more healthcare than those choosing plan L.

	Full Sample			First Time Enrollees	
				excluding zeros	including zeros
	(1)	(2)	(3)	(4)	(5)
ln expected consumption (ACG)	0.124*** (0.00517)	0.132*** (0.00519)	0.132*** (0.00513)		
ln actual consumption				0.0529*** (0.00343)	0.0649*** (0.00374)
age 30-39	0.0472*** (0.0115)	0.0478*** (0.0114)	0.0447*** (0.0113)	0.0866*** (0.0133)	0.0767*** (0.0120)
age 40-49	0.0579*** (0.0122)	0.0526*** (0.0123)	0.0502*** (0.0122)	0.115*** (0.0152)	0.102*** (0.0139)
age 50-59	0.103*** (0.0144)	0.0974*** (0.0146)	0.0928*** (0.0146)	0.124*** (0.0181)	0.118*** (0.0170)
female	0.0158 (0.00972)	0.0230* (0.0101)	0.0293** (0.0102)	0.0121 (0.0112)	0.0111 (0.0104)
worker controls	no	yes	yes	yes	yes
state FEs	no	no	yes	yes	yes
Observations	22299	22299	22299	7012	8145

Table 1.4: Regressions of plan choice on health risk. The dependent variable is an indicator for plan H, the more comprehensive option. Observations are pooled over 2005-2007, and standard errors are clustered at the person level. Worker controls include indicators for hourly, full-time, and union status. Columns 1 through 3 include all enrollees in the main estimation sample. Columns 4 and 5 are restricted to employees in their first year of health plan enrollment, and the health risk variable is actual healthcare expenditure in the first plan year. Actual expenditure includes some zeros, so it cannot be transformed into log expenditure directly. Therefore in column 4, I drop the zeros. In column 5, I add \$50 to total claims before taking logs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Full Sample		Separately by Age Group				Full Sample
	(1)	(2)	18-29 (3)	30-39 (4)	40-49 (5)	50-59 (6)	(7)
ACG prediction	1.127*** (0.0706)	1.123*** (0.0753)	1.290*** (0.228)	0.977*** (0.108)	1.137*** (0.113)	1.134*** (0.140)	1.108*** (0.0726)
Constant	7.725 (201.0)	-107.5 (150.6)	-393.9 (318.3)	290.4 (233.6)	113.9 (310.9)	-2.785 (693.2)	-237.7 (182.6)
age 30-39		30.37 (181.7)					
age 40-49		273.8 (261.2)					
age 50-59		166.3 (344.1)					
Chose Plan H							943.7*** (240.8)
Obs	22299	22299	4570	5787	6875	5067	22299
R^2	0.150	0.150	0.238	0.168	0.097	0.169	0.151
F-test for age		p = .67					

Table 1.5: Regressions of realized healthcare consumption on predictions of healthcare consumption generated by the Johns Hopkins Adjusted Clinical Grouper (ACG). No additional controls are included. Observations are pooled over 2005-2007, and standard errors are clustered at the person level. The first column includes all enrollees in the main estimation sample. Columns 2 through 5 run the regression separately by age group. The slopes are close to one and the intercepts close to zero, indicating that the prediction is unbiased. This is true for both the full sample and among each each groups, indicating that the ACG predictor seems to do an equally good job predicting risk at all ages. Column 6 includes an indicator for enrollment in plan H in the period for which the prediction is made. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Parameter	Coefficient	SE
Age Indicators (18-29 omitted category)		
30-39	\$261	(\$71)
40-49	\$323	(\$75)
50-59	\$508	(\$92)
Female	\$178	(\$50)
Plan H Intercept	-\$955	(\$153)
Coefficient of Absolute Risk Aversion	0.000696	(0.000101)
Standard Deviation of the Error: σ_{e_H}	\$1351	(\$165)
obs	8120	

Table 1.6: Results from the expected utility plan choice model. Standard errors are calculated from the inverse of numerically-computed Hessian. The Plan H intercept varies dollar-for-dollar with the assumed price difference between plans L and H, which is not observed. The value of the intercept -\$955 was estimated assuming a price difference of \$572, which equals 20% of the average cost difference between plans. If the price difference was assumed to be \$100 higher, then the absolute value of the plan H intercept would be reduced by \$100.

Reference	Absolute Risk Aversion	Certainty Equivalent
Cohen-Einav (2007) median	$3.4 \cdot 10^{-5}$	99.7
Metrick (1995)	$6.6 \cdot 10^{-5}$	99.3
Gertner (1993)	$3.1 \cdot 10^{-4}$	97.0
Handel (2010) median	$3.7 \cdot 10^{-4}$	95.2
This Paper	$6.96 \cdot 10^{-4}$	93.5
Handel (2010) mean	$7.9 \cdot 10^{-4}$	92.6
Sydnor (2006)	$2.0 \cdot 10^{-3}$	83.3
Cohen-Einav (2007) mean	$3.1 \cdot 10^{-3}$	76.5
Holt and Laury (2002)	$3.2 \cdot 10^{-2}$	21.0

Table 1.7: Comparisons of the estimated risk coefficient to estimates from other insurance studies. To aid interpretation, the last column displays a certainty equivalent measure of the risk parameters. The certainty equivalent here is the amount X that would make someone indifferent between accepting a gamble in which they win \$100 or lose \$ X with equal probability versus a status quo where nothing happens.

**Welfare changes under counterfactual prices:
Relative to uniform pricing baseline with 80% subsidy**

	P(H)-P(L)		Fraction in H		Welfare Gain		obs
	Old ΔP	New ΔP	Old	New	\$ Change	% Change	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Optimal Uniform Price							
Overall	\$688	\$390	0.32	0.40	\$11.12	0.33%	8,084
18-29	\$688	\$390	0.19	0.26	\$16.78	0.50%	1,683
30-39	\$688	\$390	0.29	0.37	\$16.09	0.48%	2,103
40-49	\$688	\$390	0.35	0.43	\$10.05	0.30%	2,521
50-59	\$688	\$390	0.44	0.53	\$1.39	0.04%	1,777
Panel B: Optimal Age-specific Prices							
Overall	688	—	0.32	0.40	\$12.12	0.36%	8,084
18-29	\$688	\$285	0.19	0.28	\$18.13	0.54%	1,683
30-39	\$688	\$325	0.29	0.39	\$16.56	0.49%	2,103
40-49	\$688	\$410	0.35	0.43	\$10.23	0.30%	2,521
50-59	\$688	\$525	0.44	0.49	\$3.84	0.11%	1,777
Panel C: Minimal Information Age-specific Prices							
Overall	\$688	—	0.32	0.43	\$10.55	0.31%	8,084
18-29	\$688	\$95	0.19	0.33	\$14.06	0.42%	1,683
30-39	\$688	\$255	0.29	0.40	\$15.98	0.47%	2,103
40-49	\$688	\$335	0.35	0.45	\$9.36	0.28%	2,521
50-59	\$688	\$433	0.44	0.52	\$2.49	0.07%	1,777

Table 1.8: Pricing and welfare changes associated with three counterfactual pricing policies (discussed in the text). In the common baseline, the employer is assumed to set uniform prices for each plan, and to set those prices equal to 20% of the average plan costs. The first two columns report how the relative price of plan H changes under the counterfactual. Columns (3) and (4) report the old and new take up rates for plan H. Column 5 reports changes to welfare (total surplus) per enrollee per year in dollars. Column 6 scales these changes by average total healthcare expenditure over the entire sample.

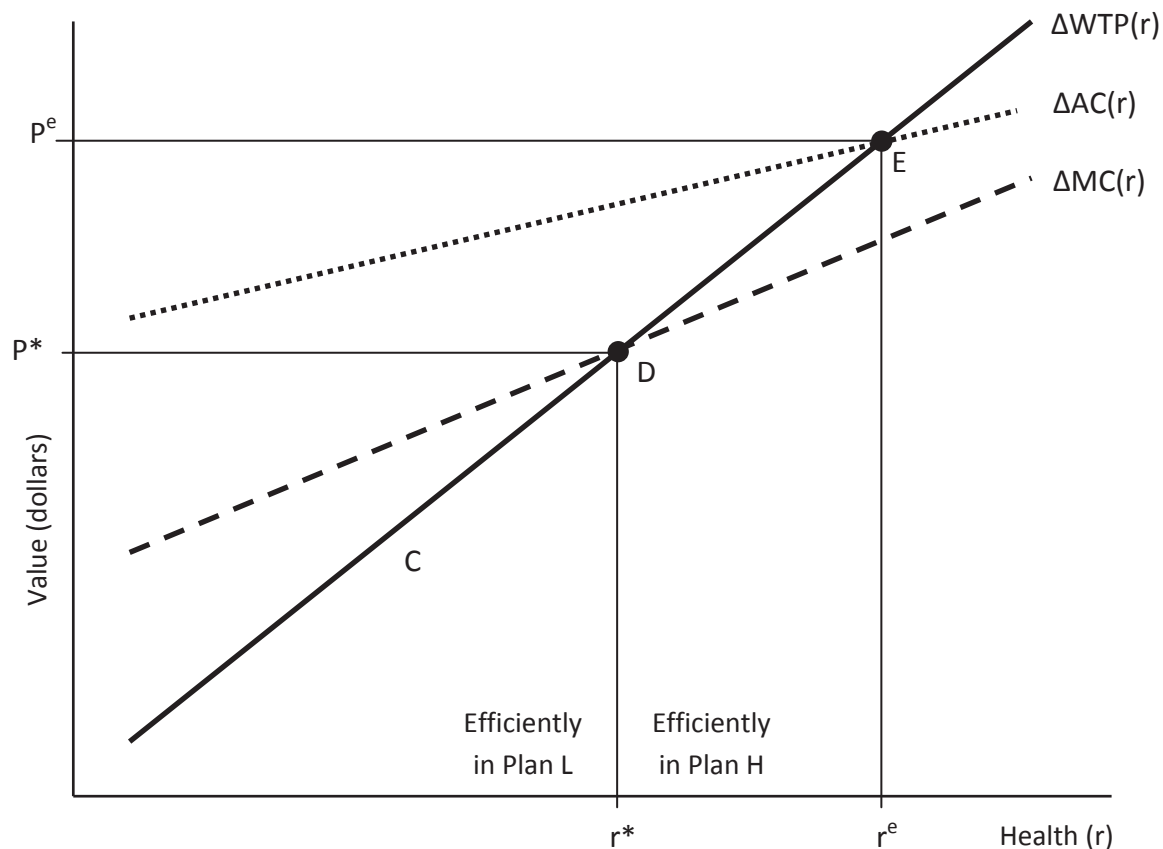


Figure 1.1: Equilibrium and efficient insurance allocations. Efficient allocations occur at point D, where the marginal cost and willingness-to-pay curves cross. Agents with $r < r^*$ efficiently choose plan L, and those with $r \geq r^*$ efficiently choose plan H. Competitive equilibrium allocations occur at point E, where $\Delta AC(r)$ intersects the willingness-to-pay curve. In the employer setting, the equilibrium allocation may lie to the left or right of D, depending on how the employer chooses to subsidize average costs in each plan.

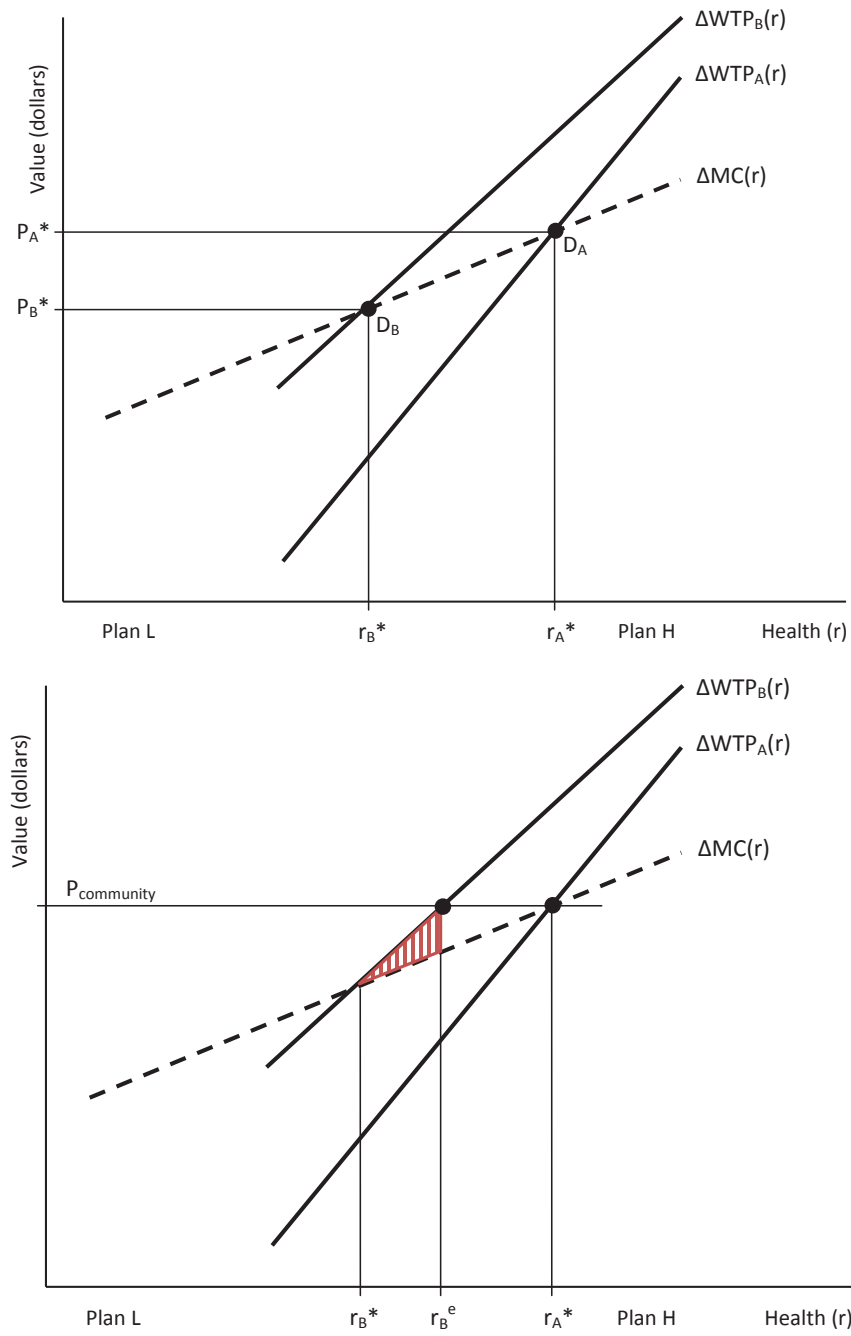


Figure 1.2: No single price can create efficient insurance allocations when there is heterogeneity in willingness to pay. The efficient price for group A occurs at P_A^* , which induces all group A types with health risk exceeding r_A^* to choose plan H. The efficient price for group B occurs at P_B^* , which induces the risk cutoff r_B^* . The bottom panel highlights that any community price that induces efficient sorting for one group induces inefficient sorting for the other.

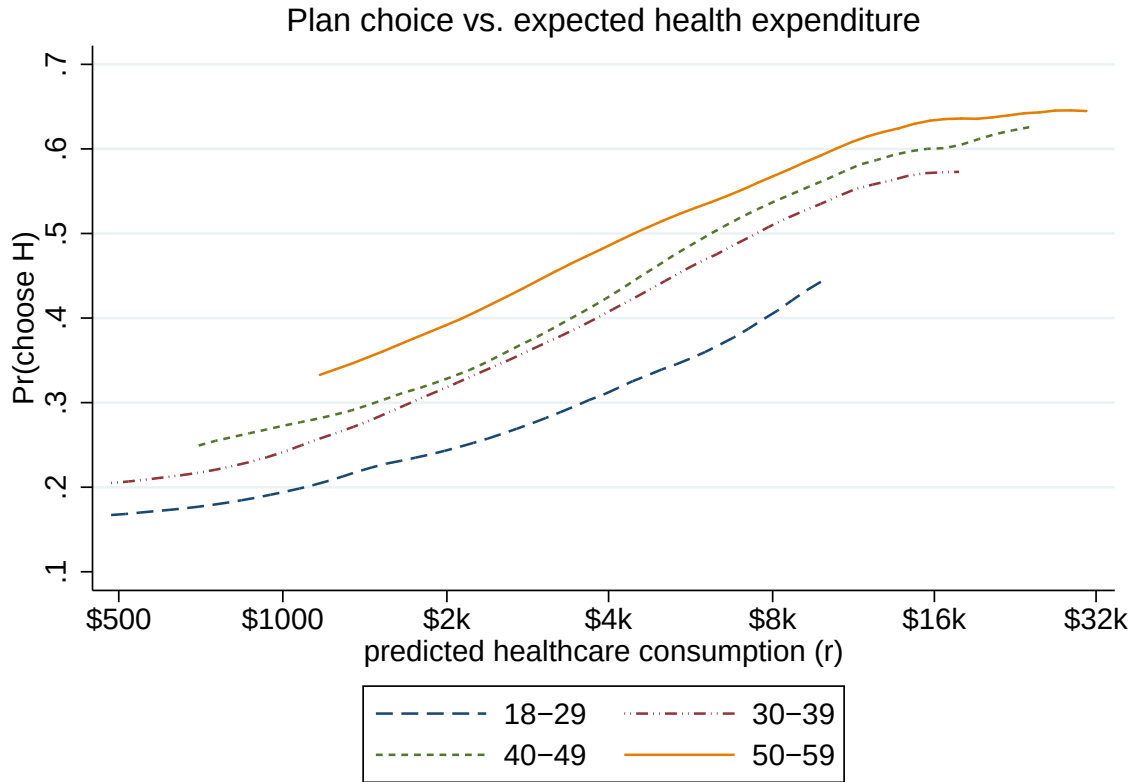


Figure 1.3: Local linear regressions of plan choice versus expected healthcare consumption. For any fixed level of expected consumption, the fraction choosing plan H in any age group is just equal to the fraction of the group for whom willingness to pay exceeds price. The diagram reveals that willingness to pay is monotonically increasing in age.

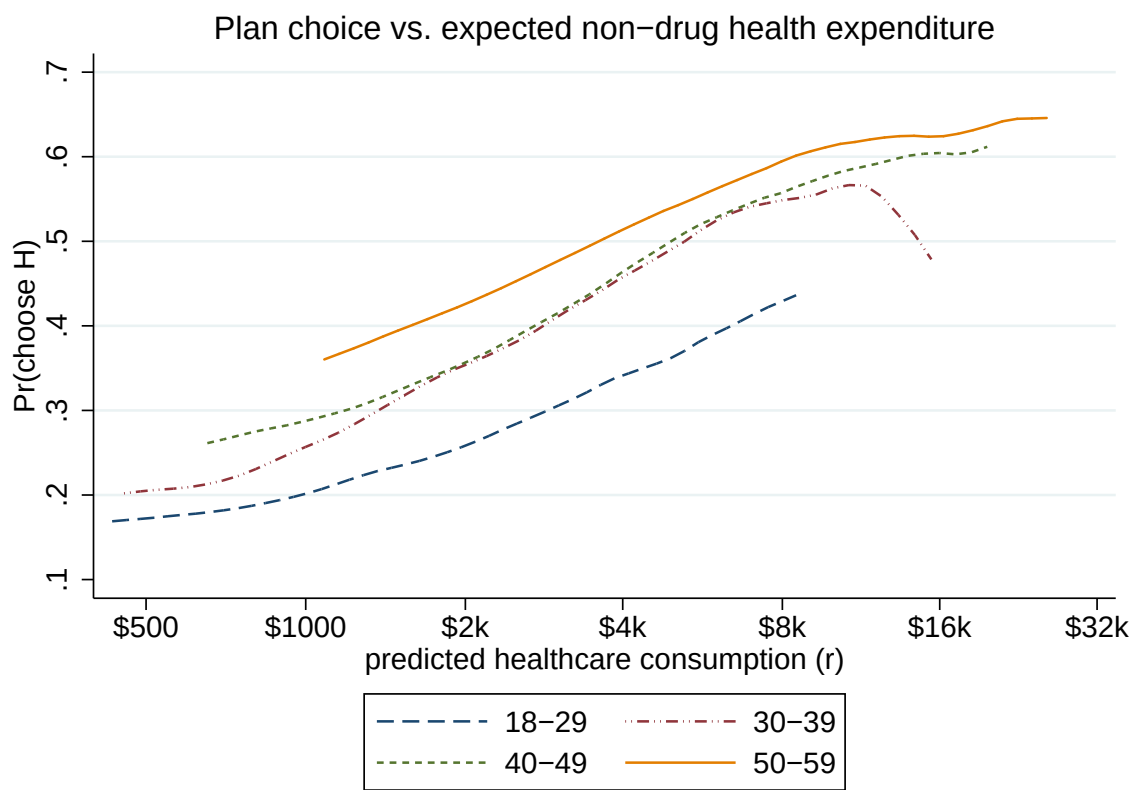


Figure 1.4: Local linear regressions of plan choice versus expected healthcare consumption, excluding predicted expenditure on prescription drugs. For any fixed level of expected consumption, the fraction choosing plan H in any age group is just equal to the fraction of the group for whom willingness to pay exceeds price. The diagram reveals that willingness to pay is monotonically increasing in age.

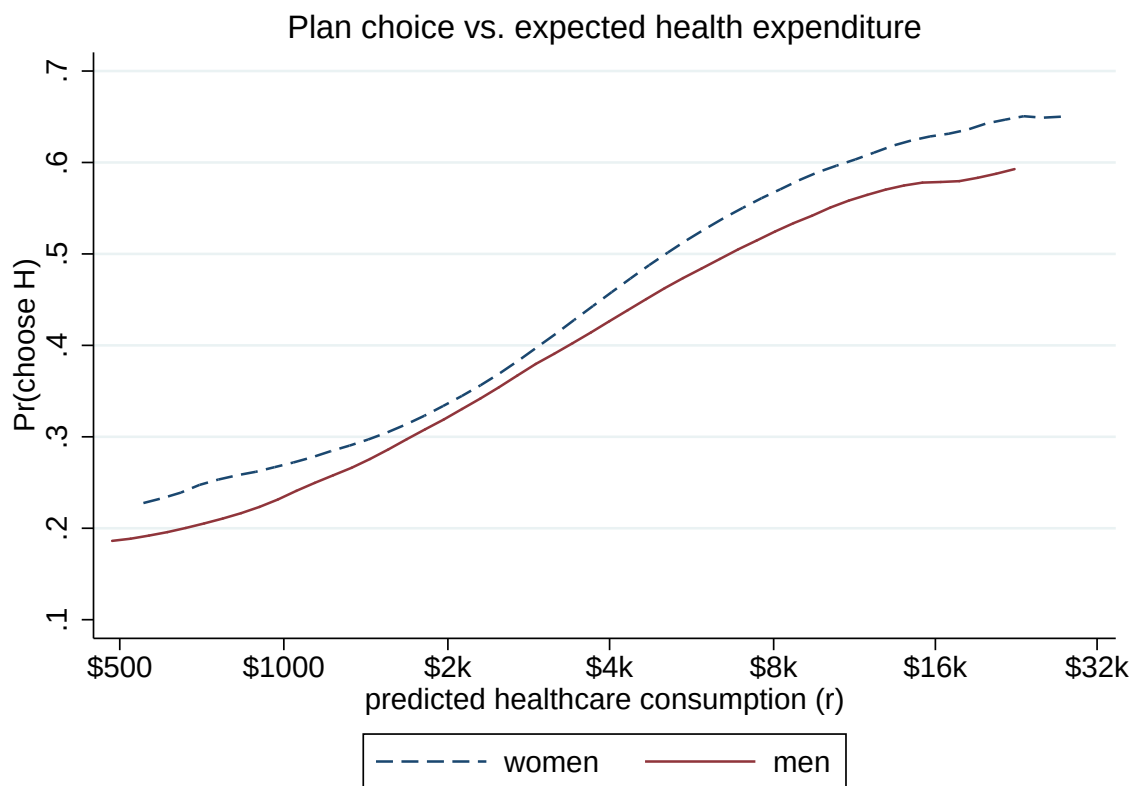


Figure 1.5: Local linear regressions of plan choice versus expected healthcare consumption by gender. For any fixed level of expected consumption, the fraction choosing plan H in any age group is just equal to the fraction of the group for whom willingness to pay exceeds price. The diagram reveals that willingness to pay seems slightly higher for women than for men.

Local Polynomials: Plan Choice on Last-Year's Expenditure Heuristic 2005–2007

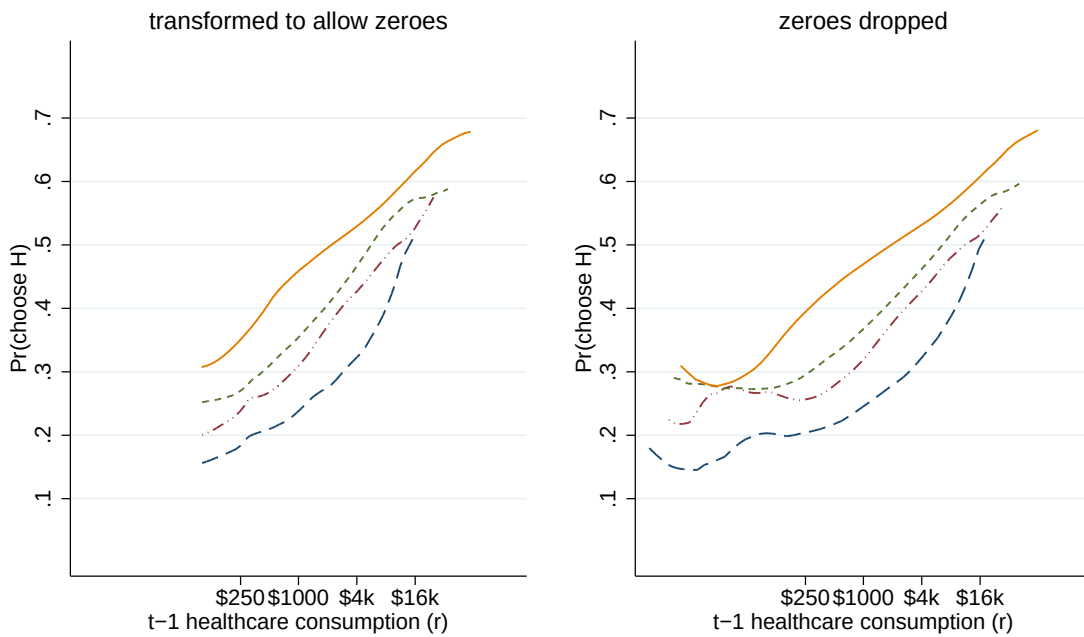


Figure 1.6: Local linear regressions of plan choice versus last period's health care consumption. As in earlier figures, the four lines correspond to the four age groups: 18-29, 30-39, 40-49, and 50-59. The diagram reveals that willingness to pay is monotonically increasing in age at any level of health care consumption.

Local Polynomials: Plan Choice on Realized Expenditure (Ex-post) 2005–2007

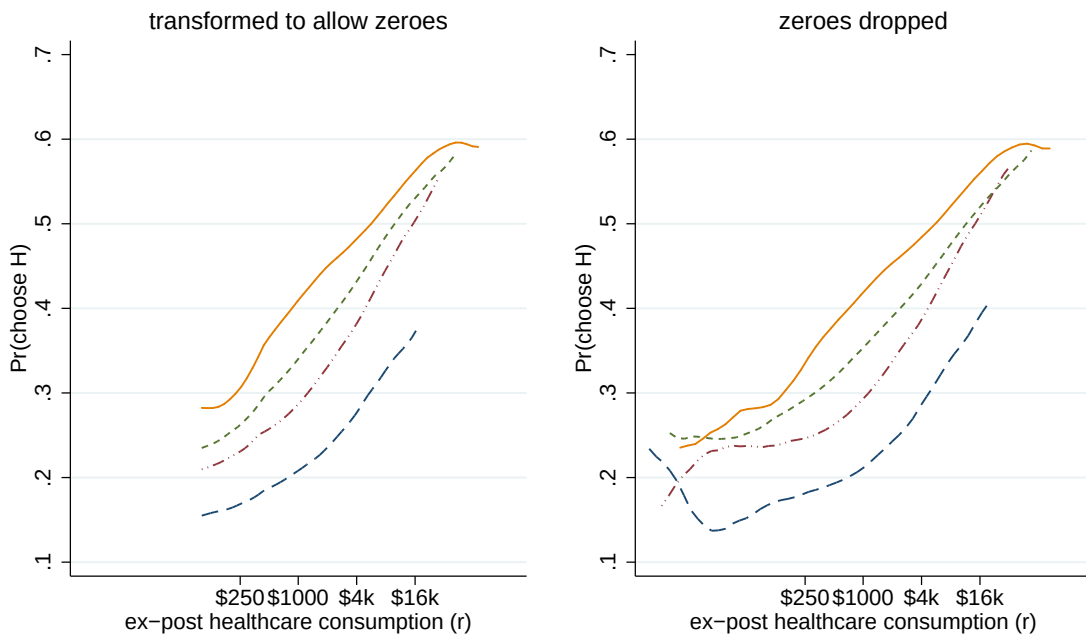


Figure 1.7: Local linear regressions of plan choice versus ex-post health care expenditure in the period for which the plan choice was made. As in earlier figures, the four lines correspond to the four age groups: 18-29, 30-39, 40-49, and 50-59. The diagram reveals that willingness to pay is monotonically increasing in age at any level of health care consumption.

Chapter 2

Black-White Disparities in Life Expectancy: How Much Can the Standard SES Variables Explain?

2.1 Introduction

In the United States, whites live longer than blacks. Since 1980, the difference in life expectancy at birth between whites and blacks has ranged from 5.1 to 7.1 years (Arias et al. 2010).¹ Mortality differences across racial lines have attracted significant public policy interest in the past two decades, which has been met with a surge of scholarly research on racial disparities in health (for an overview, see Harris 2010). Much of this scholarly work has focused on the role of socioeconomic status (SES) in explaining racial differences in health outcomes, and numerous studies have attempted to decompose racial mortality differences, usually between blacks and whites, into a component that is attributable to group differences in SES. One limitation of past studies is that by focusing on mortality hazards estimated over limited age ranges, these studies have left largely unexamined the extent to which SES explains disparities in life expectancy.

This article quantifies the proportion of the racial gap in life expectancy that can be accounted for by differences in characteristics such as income, education, employment and occupation, homeownership, urban residence, and marital status. It employs a reweighting

¹This article uses data supplied by the National Heart, Lung, and Blood Institute; NIH; and DHHS from the National Longitudinal Mortality Study. The views expressed in this article are those of the author and do not necessarily reflect the views of the National Heart, Lung, and Blood Institute; the Census Bureau; or the National Center for Health Statistics.

technique that facilitates the decomposition of mortality differences into a component that reflects compositional differences in socioeconomic and demographic characteristics between blacks and whites.

The technique is compatible with a focus on age-specific mortality rates, excess mortality, or any other measure of mortality, but because the data used here cover almost the entire age distribution, it affords a unique opportunity to investigate how SES differences can account for mortality differences over the life course, as summarized by life expectancy. This is the first study able to offer a precise quantitative result regarding the proportion of the black-white difference in life expectancy that can be accounted for by differences in individual-level SES variables. Using data from the National Longitudinal Mortality Study (NLMS)—a large, nationally-representative data set with rich information on SES—I find that for males, about 80% of the black-white gap in life expectancy at age 1 can be accounted for by differences in socioeconomic and demographic characteristics; 70% of the gap between black and white females is explained. I also find that labor force participation; occupation in a white-collar job; and, among women only, marital status have no power to explain the black-white life expectancy gap after precise measures for income and education are controlled for.

The hypothesis that racial differences in socioeconomic characteristics can account for some or all of the mortality differences between blacks and whites is quite natural given that income (Rogot et al. 1992; Sorlie et al. 1995; Elo and Preston 1996; Ecob and Smith 1999; Case et al. 2002; Winkleby and Cubbin 2003; Warner and Hayward 2006; Case and Paxson 2011); wealth (Warner and Hayward 2006), education (Rogot et al. 1992; Sorlie et al. 1995; Elo and Preston 1996; Winkleby and Cubbin 2003; Warner and Hayward 2006; Kimbro et al. 2008; Meara et al. 2008); employment and occupation (Rogot et al. 1992; Johnson et al. 1999; Sorlie et al. 1995; Winkleby and Cubbin 2003; Warner and Hayward 2006); marital status (Hu and Goldman 1990; Sorlie et al. 1995; Elo and Preston 1996; Johnson et al. 2000; Warner and Hayward 2006); and, to a lesser extent, urban residence (Elo

and Preston 1996; Hayward et al. 1997) have each been shown to be predictive of health outcomes, and blacks systematically attain levels of these variables that would imply a health disadvantage relative to whites.

In two widely cited reviews, Williams and Collins (1995) and Hummer (1996) summarized the literature on race, SES, and health as indicating that racial differences in socioeconomic characteristics can explain much of the black-white mortality gap. The individual studies cited in these reviews (e.g., Rogers 1992), as well as the more recent studies that have followed them (e.g., Elo and Preston 1996; Hayward et al. 1997; Howard et al. 2000; Richardus and Kunst 2001; Winkleby and Cubbin 2003; Warner and Hayward 2006), generally estimate hazards of dying that include an indicator for race. Accounting for the role of SES is accomplished by comparing the magnitude and significance of the race coefficient before and after the inclusion of SES variables. These studies generally estimate hazards over a limited age range, and sometimes for only particular causes of death. The focus of these analyses is justified both by the specific aims of the research and by data limitations because many data sets cover only adult deaths in some age range. However, a natural consequence of this narrow focus is that the interested reader would find it difficult or impossible to derive from the existing body of research an answer to the question of how much of the black-white gap in life expectancy could be attributable to group differences in socioeconomic and demographic characteristics. This is in spite of the fact that almost all studies examining racial mortality differences have motivated their analyses by pointing to racial gaps in life expectancies.

The intuition behind the decomposition methodology used here is to reweight individual respondents in the black subsample according to their SES characteristics so that in aggregate, the reweighted black sample matches the original white sample in terms of relevant characteristics. To take a simple example, if the proportion of low-income respondents in the white sample is 33% and the proportion of low-income respondents in the black sample is 50%, then the reweight would scale down the contribution of low-income blacks to

black mortality rates by a factor of $2/3$ and scale up the contribution of high-income blacks by a factor of $4/3$. This would result in a “counterfactual” black sample that resembles the actual white sample in that 33% of its members were low income. Mortality in the counterfactual black sample can then be compared with the actual white mortality.

The idea of decomposing compositional differences between groups by reweighting is not new (see, e.g., DiNardo et al. 1996). This is, however, the first time that this decomposition strategy has been applied to racial differences in life expectancy. The technique is useful because life expectancy is not an individual-level outcome; therefore, it is not possible to treat it as a left hand-side variable in a regression on individual-level characteristics.²

Throughout the analysis, I mechanically treat the SES and demographic variables as explanatory factors for life expectancy. This approach follows much of the literature, but I view it as generating results that form an upper bound on the causal importance of these variables in explaining mortality because it is likely that causality works both from SES to health and from health to SES (Deaton 2002). Although investigation into the causal mechanisms working from SES to health are of paramount importance for developing policy responses to racial disparities, the focus here is not on mechanisms but rather on a careful statistical accounting. I return to this issue in the Discussion section.

The remainder of this article is organized as follows. The next section describes in detail the NLMS. The Methods section explains and formalizes the reweighting technique. Finally, the main results, which take the form of counterfactual black life expectancies, are reported and discussed, and several robustness checks are presented.

²Potter (1991) regressed racial differences in life expectancy across metropolitan statistical areas (MSAs) on MSA characteristics. Importantly, whereas Potter explored the effect of area-level socioeconomic characteristics, such as the local employment and poverty rates, the data and technique used here permit attribution of life expectancy differentials to *individual*-level characteristics, including own employment status and own income.

2.2 Data

This article uses data from Release 3 of the NLMS public-use file. The NLMS is a prospective mortality study of the noninstitutionalized U.S. population. It links demographic and socioeconomic data from Current Population Surveys (CPS) with death records. The CPS, which forms the baseline of the study, is a nationally representative household survey administered monthly by the U.S. Census Bureau and the Bureau of Labor Statistics. The emphasis of the CPS is on labor force participation and employment, but the survey also collects data on income, education, and a variety of household demographics. Mortality follow-up for 11 years following each baseline CPS is accomplished by matching CPS records to death certificates through the National Death Index (NDI), which is an electronic census of all U.S. death certificates since 1979. Eleven CPS cohorts from the early 1980s constitute the NLMS public-use file. These cohorts are aggregated, and balancing weights are applied so that the aggregated sample reflects the noninstitutionalized U.S. population as of April 1, 1983.

In this study, I compare mortality by sex between non-Hispanic whites and non-Hispanic blacks. From the pool of white and black respondents with a valid survey weight, I eliminate records with missing household income or with missing classification as urban/non-urban. I also eliminate records for individuals aged 25 and older with missing marital status, working status, or educational attainment. These restrictions reduce the sample size by less than 0.02%. The resulting sample covers 869,084 individuals and more than 9 million life-years of follow-up. Summary statistics by race and sex at the individual level are provided in Table [2.1](#).

The summary statistics show that incomes are lower among blacks than among whites. Black households are larger, which magnifies differences in household income. Black respondents are less likely to own their residence and more likely to live in an urban neighborhood. Homeownership is reported as a potential proxy for wealth because no direct measure of

wealth exists in the NLMS.

Among adults aged 25 and older, blacks have lower educational achievement; in particular, white men in the sample are 2.4 times as likely as black men to earn a college degree or higher. Black men (but not women) are less likely to be in the labor force compared with whites; and conditional on labor force participation, black men and women are both somewhat less likely to be employed and much less likely to be working in a white-collar occupation. Marriage rates among black respondents are significantly lower as well, especially for black women.³

In the decomposition that follows, the variables used to calculate the reweighting functions include family income, education, current marital status, homeownership, labor force participation, white-collar occupation, and urban residence. Family income is deflated to 1990 dollars and then reexpressed as a percentage of the 1990 poverty line, where the poverty line is specific to a particular household makeup. This allows the income measure to account for differences in household composition, which is important in this context because blacks are more likely to live in larger households. This relative-to-poverty-line income, which is coded categorically as percentiles in the NLMS data set, enters the analysis as nine income levels, roughly corresponding to deciles. Education is categorized in the same six groups as reported in Table 2.1. Homeownership, urban residence, current marital status, and labor force participation are all binary variables. White-collar status takes three values: white collar, blue collar, and missing (to mechanically allow for respondents who are out of the labor force). Family income, homeownership, and urban residence are common to all family members in the household survey and are included in the reweight for all age groups. Education, marital status, and labor market activities are individual attributes included in the reweight for adults aged 25 and older.

An important limitation of the NLMS is that neonatal death is largely missed. Be-

³The sex difference in current marital status is due entirely to older respondents. Because women live longer than men, they are more likely to be (unmarried) widows and skew the sex balance of marriage rates among both whites and blacks.

cause the baseline survey is derived from the household roster of the CPS, children dying within a short window after birth, possibly in the hospital, have a low probability of being enumerated at baseline. This could lead to downward-biased estimates of mortality at age 0. For this reason, in all the analyses that follow, I present results on life expectancy at age 1 rather than age 0. I discuss the likely consequences of ignoring infant mortality in the Discussion section.

Because the object of interest in this article is a comparison of life expectancy across race, it is important to confirm that life expectancy estimates derived from NLMS replicate the racial disparities found in the standard estimates of life expectancy, such as those published by the National Center for Health Statistics (NCHS). There are several reasons that estimates from the NLMS and NCHS should not match exactly. For one, the NLMS covers only the noninstitutionalized U.S. population. More importantly, the NLMS covers the mortality experience of CPS cohorts spanning a decade, each with a mortality follow-up of 11 years. Pooling across CPS cohorts is useful for creating a sample large enough to support the analysis, but the pooled estimates will not be directly comparable with any yearly estimate from the NCHS.

Nonetheless, to investigate whether the NLMS captures the broad patterns of racial disparities evident in other national data, Table 2.2 lists my estimates of life expectancy at age 1 (e_1) derived from the NLMS sample and compares these with estimates from the 1987 and 1988 NCHS (National Center for Health Statistics a,b).

Estimates of e_1 in the NLMS are calculated throughout this article according to standard life table procedures but without making functional form assumptions on the survival curve. Age-specific mortality rates are estimated in one-year intervals up to age 85. For the open-ended age interval beginning at 85, calculations follow standard practice (Preston et al. 2001, p. 48). Because the NLMS public-use file is composed of CPS cohorts from the early 1980s with a mortality follow-up period of 11 years, I include NCHS estimates from 1987 and 1988 for comparison. Of course, there is no reason to expect the pooled NLMS

estimates to exactly match the single-year NCHS estimates, but patterns of mortality by race and sex found in the NLMS are broadly consistent with the NCHS.⁴ The racial mortality gap is slightly smaller in the NLMS estimates, which estimate black life expectancy at age 1 to be higher than in the NCHS estimates. However, the NCHS numbers reflect the uncommonly bad mortality experience among blacks in the late 1980s, which generated period life expectancies among blacks lower than in the years both preceding and following (Arias et al. 2010).

2.3 Methods

2.3.1 Decomposition Strategy

This reweighting technique used in this article is conceptually related to the decompositions introduced by Blinder (1973) and Oaxaca (1973) to evaluate whether differences in group characteristics, such as education and work experience, could account for the observed differences in wages between men and women or whites and blacks. The Blinder-Oaxaca decompositions relied on the assumption of a linear functional relationship between the independent and dependent variables at the individual level. A direct application of the Blinder-Oaxaca decomposition would be impossible in this context because e_1 is not an individual-level outcome, but it is straightforward to construct a generalized decomposition in the same spirit as Blinder-Oaxaca. Denote white and black life expectancy at age 1 by e_1^{white} and e_1^{black} , respectively. By adding and subtracting a term, one can write the difference in life expectancy as

$$e_1^{\text{white}} - e_1^{\text{black}} = (e_1^{\text{white}} - e_1^{\text{black,CF}}) + (e_1^{\text{black,CF}} - e_1^{\text{black}}) \quad (2.1)$$

⁴Because the NLMS excludes institutionalized persons, it is somewhat surprising that the NLMS estimates of e_1 for whites fall slightly below the NCHS figures. For most ages, age-specific mortality is indeed lower in the NLMS. The lower life expectancy among whites in the NLMS is driven entirely by the age ranges 60 to 85 for men and 65 to 85 for women, over which the white NLMS cohorts experience higher mortality compared to the NCHS.

where $e_1^{\text{black,CF}}$ is a counterfactual life expectancy among a sample of blacks with characteristics that match the white sample. The second difference contained in parentheses in (2.1) is often referred to as the part of the difference in the outcome (here $e_1^{\text{white}} - e_1^{\text{black}}$) that is due to differences in characteristics. Analyzing the fraction of the total gap in life expectancy that is due to differences in characteristics is the focus of this article.

Using the demographic and socioeconomic variables from the baseline CPS surveys, I calculate a variety of counterfactuals in which the black subsample is reweighted to match the white subsample along one or several dimensions of observable characteristics. The details of constructing are presented in the next subsection. To quantify the proportion of the black-white mortality gap that can be accounted for by differences in characteristics, I report the counterfactual life expectancies as well as the fraction of the gap explained:

$$\% \text{ Explained} = \frac{e_1^{\text{black,CF}} - e_1^{\text{black}}}{e_1^{\text{white}} - e_1^{\text{black}}}, \quad (2.2)$$

where the numerator measures difference between the counterfactual and true life expectancies for blacks, and the denominator measures the difference in (true) life expectancy by race. True life expectancies in Eq. (2) are calculated as those reported in Table 2.3. *% Explained* is intended in the accounting sense, rather than as a strong claim about causality.

2.3.2 Generating Reweighting Functions and Constructing Counterfactuals

Constructing counterfactual life expectancies ($e_1^{\text{black,CF}}$) here involves introducing a reweighting term that multiplies each observation's contribution to the calculation of age-specific mortality rates (ASMRs). Denote the probability density of observable characteristics in each of the black and white subsamples as $f(y|R=b)$ and $f(y|R=w)$, respectively, where R specifies race; and y is a vector of observables, such as income, education, and marital status. In this application, f is always a probability mass function, rather than a continuous density, because the NLMS covariates are coded discretely. Nonetheless, the results

are general to both discrete and continuous covariates. The reweighting function $\psi(y)$ is defined as a function that multiplies $f(y|R=b)$ so that the product equals $f(y|R=w)$ at each value y . In other words, $\psi(y)$ forces the distribution of characteristics to match some reference distribution: in this case, whites. For illustration, I will derive $\psi(y)$ for the case of matching the income distribution of blacks to whites. Additional reweighting functions are derived in the same manner, and the formulas for each are explicitly listed in the tabulations of results.

Let i and a represent income and age, and let y' represent the elements of y other than i and a . Consider the counterfactual in which the black income distribution matches the white income distribution at each age, but the joint conditional distribution of other attributes $f(y'|i, a, R=b)$ remains the same as in the original black sample. The counterfactual density of attributes in this case can be written as the product of several observable conditional density terms: $f(y'|i, a, R=b)f(i|a, R=w)f(a|R=b)$. It is useful to construct a reweighting function as follows:

$$f(y'|i, a, R=b)f(i|a, R=w)f(a|R=b) \tag{2.3}$$

$$\begin{aligned} &= f(y'|i, a, R=b)\psi_{i|a}(i, a)f(i|a, R=b)f(a|R=b) \\ &= f(i, a, y'|R=b)\psi_{i|a}(i, a) \end{aligned} \tag{2.4}$$

where the reweighting function $\psi_{i|a}(i, a)$ is defined as

$$\psi_{i|a}(i, a) \equiv \frac{f(i|a, R=w)}{f(i|a, R=b)}, \tag{2.5}$$

(2.4) shows that the desired counterfactual can be achieved by simply multiplying the true distribution of characteristics by an easily calculable reweighting function. The reweighting function, (2.5), is just the density of income (conditional on age) among whites

divided by the same density among blacks. Given the income distribution shown in Table 2.1, this implies that for high levels of income, the numerator in (2.5) will be larger than the denominator because relatively more whites than blacks are found at higher income levels. This results in a reweighting function that is greater than 1 at high incomes, up-weighting observations of wealthier blacks within each age band.

This reweighting function can be directly applied to the calculation of ASMRs in order to generate counterfactual mortality rates and, in turn, a counterfactual life expectancy. The counterfactual ASMRs, estimated in one-year bins and denoted ${}_1m_x$ are:

$${}_1m_x = \frac{\sum_{j \in J} D_j(x) \cdot wt_j \cdot \psi_{i|a}(i, a)}{\sum_{j \in J} L_j(x) \cdot wt_j \cdot \psi_{i|a}(i, a)}, \quad (2.6)$$

where subscript j indexes individuals; $L_j(x)$ and $D_j(x)$ are defined as in life tables; $D_j(x)$ serves as an indicator for death at age-last-birthday x during the follow-up period; and $L_j(x)$ represents the fraction of a year, if any, lived in the age interval $(x, x+1)$ during the follow-up period. The weight, wt_j , is the sample weight original to the NLMS data set. Except for the reweighting function $\psi(\cdot)$, (2.6) is just a standard formula for an ASMR.

For every counterfactual considered, reweighting functions are specific to a five-year age group. This age grouping has nothing to do with the calculation of age-specific mortality rates, which are done in one-year bins. Instead, this grouping is necessary to ensure, for example, that a wealthy younger black respondent is not up-weighted to balance a wealthy older white respondent.

Calculating the reweighting functions in practice requires that the actual probability mass function for blacks includes full support over the range of characteristics for which the white probability mass function has support. In other words, to reweight observations in the black subsample that possess some specific combination of socioeconomic attributes, it must be the case that at least some black respondents with that particular combination

of attributes are observed. For a few (of the thousands) of cells defined by unique combinations of income, education, marital status, and other attributes, no black respondents are observed. Mathematically, such empty cells would correspond to a 0 in the denominator of (2.5).

To address this difficulty, in the analysis that follows, I drop white observations for which no black match is observed—about 3% of the white sample.⁵ This creates a selected sample of whites and slightly lowers life expectancy in the restricted white sample. Among white men, the estimate of life expectancy at age 1 drops from 71.3 to 71.2 after removing such cases. Among white women, it declines from 78.1 to 77.9.

To keep comparisons consistent, in the analysis that follows, I always compare the original and counterfactual black life expectancies with these revised white estimates. An alternative approach to addressing the lack of joint support in the probability mass functions of characteristics would be to make the grid of characteristics less fine, such as by lumping more education or income groups together. However, to do so would be to impose more parametric assumptions on the functional relationship between socioeconomic variables and mortality, which would affect results in a far less transparent way than dropping some white observations and reporting the effect on white life expectancy. Nonetheless, as a robustness check I also report results of an alternative scheme that reweights whites rather than blacks. That method drops almost no observations and yields results that closely mirror the main results.

2.3.3 Relationship to Other Demographic Decompositions

An important class of demographic techniques, formalized by Kitagawa (1955), decomposes differences in crude (or total) rates between groups into a component that reflects a compositional difference and a component that reflects differences in specific rates. The

⁵The reverse scenario, of no support in the white density function but positive support in the black density function, poses no technical problem. It results in a reweighting term equal to 0.

prototypical example in Kitagawa (1955) is the decomposition of a difference in crude death rates into a part attributable to differences in the age distributions of the populations and a part attributable to differences in age-specific death rates. Variants on the technique differ in technical details but share the aim of purging or measuring the compositional effects.

The decomposition here is similar in that it evaluates a certain type of compositional effect in accounting for differences in group mortality. Here, the compositional effect to be measured and removed is that of SES differences between race groups. Just as with most standardization techniques, the reweight here permits any distribution (of SES variables) to be used as the “standard” or “reference” distribution. The analysis that follows primarily uses the white sample as the reference population, but also presents results in which the black sample is used as the reference.

Compared with Kitagawa (1955) and later work that builds on it, including Das Gupta (1993), there are two important differences in the decomposition technique described here. The first is that this reweight is applied to individual observations rather than to rates or other statistics.

A second difference and a subtle advantage of the reweighting technique presented here is that the counterfactual generated by an individual-level reweight can be used to calculate any group statistic, including the median and variance, and is not limited to the standardization or decomposition of means, as in Kitagawa (1955), Das Gupta (1993), and the Blinder-Oaxaca method. For a more technical exposition of the reweight details, see DiNardo et al. (1996).

A separate class of demographic decompositions operates on life table quantities to attribute differences in life expectancy between two populations to differences in age-specific mortality rates (Arriaga 1984) or to differences in cause-specific mortality rates (Arriaga 1989). This article is related to, but distinct from, the strand of research on the black-white mortality gap that has focused on such mechanical decompositions. Keith and Smith (1988) decomposed racial differences in life expectancy by cause of death to identify the

leading cause contributors. More recently, Harper et al. (2007) reported on the age groups and causes of death that contributed most substantially to changes in the black-white life expectancy gap during the 1980s and 1990s.

Reweighting is a natural complement to such decompositions. After counterfactual black or white samples are created through the individual-level reweight and counterfactual life tables are created, any standard life table technique can be employed. For instance, as a supplementary analysis, I follow Arriaga (1984) to evaluate which age ranges are mechanically responsible for the increase in life expectancy after the black sample has been purged of compositional effects of SES via the reweight. Thus, the present reweighting methodology augments long-established life table decomposition techniques.

2.4 Results

2.4.1 Main Results

This section quantifies the extent to which controlling for socioeconomic and demographic covariates via a reweighting function can close the racial mortality gap, measured here as life expectancy at age 1. Tables 2.3 and 2.4 report the main results of the article for men and women, respectively. Column 1 in each table presents the sample estimate of life expectancy at age 1. This column contains the true sample estimate of white and black life expectancy in the first two rows, followed by various counterfactual estimates of black life expectancy. The set of covariates on which the reweight is performed are displayed to the left. Standard errors are in column 2. The third column computes the level difference of life expectancy between the white sample and the counterfactual black sample, and the fourth column displays the percentage of the gap ($e_1^{\text{white}} - e_1^{\text{black}}$) explained, defined as in (2.2). Here, 0% explained would correspond to a black life expectancy that did not change from its baseline after reweighting the black sample, whereas 100% explained would mean that black life expectancy converged to white life expectancy after matching characteristics.

The final column displays the reweighting function numerator for reference.

Because the reweighting methodology used here does not allow the predictive power of a set of covariates to be divided into the contribution of each covariate separately, Tables 2.3 and 2.4 present a large set of counterfactual black life expectancies in order to illuminate the relative importance of each covariate. The first set of reweights creates counterfactuals that match on only a single characteristic. Then multiple characteristics are included simultaneously so that the black sample is made to match the distribution of characteristics in the white sample along two, three, or four dimensions. By scanning down the table, the reader can judge the relative importance of various characteristics by observing how the inclusion of an additional reweighting variable affects the total predictive power of the set. For instance, among men, income alone accounts for 52% of the gap, and education alone accounts for 26%, but income and education (which are positively correlated) together account for 62% of the racial gap in life expectancy.

For both men and women, income is the most important sole predictor of life expectancy, accounting for 52% and 59% of the gap among males and females, respectively. Education and marital status are also important standalone predictors among both sexes, ranging in predictive power from 20% to 29%. To a lesser extent, homeownership (which may be serving as a proxy for unobserved wealth) is predictive, explaining 16% and 23% among males and females, respectively. Urban residence seems to have almost no predictive power by itself, and men and women differ in the extent to which labor market activities can account for any part of the gap. Recall that unlike black men, black women are more likely than their white counterparts to be participating in the labor force, so it is unsurprising that the effect of reweighting labor market activity would not be symmetrical across gender. Moreover, labor market activity and occupation covary strongly with income and education, making it most appropriate to consider the contribution of these characteristics jointly with income, and perhaps with education. Although reweights for single variables are useful for gauging the relative importance of each variable in a set, it is most informa-

tive to consider the maximum predictive power when multiple dimensions of characteristics are matched simultaneously, as is done toward the bottom of each table.

One striking difference between the results for males and females is the contribution of marital status, after income and education are controlled for. Among men, additionally controlling for marital status after controlling for income and education yields a 13 percentage point ($75\% - 62\%$) increase in the joint predictive power of the variables. Among women, the additional predictive power of marital status conditional on income and education is 0. This does not imply that marriage is an unimportant explanatory factor, given that its individual contribution in a reweight on income, education, and marital status cannot be separately identified.

There are two reasons, in principle, that marital status could have no additional predictive power among women in this case: either there is no mortality gradient in marital status among black women conditional on income and education, or the process of reweighting black women to match white women along education and income creates a reweighted black sample that incidentally also matches the white sample in terms of the proportion who are married so that there is no compositional difference in marriage left to remove. Here, a lack of a mortality gradient in marriage (conditional on income and education) is, in fact, driving the result because even after reweighting for income and education, marriage rates among black women are still 15 percentage points lower than among white women. The difference across gender is consistent with Hu and Goldman (1990), who found a stronger effect of marriage for men than women and with Waldron et al. (1996), who found that the health benefits of marriage among women are stronger for white than black women and operate partly through an income channel.⁶

Jointly controlling for income, education, and marital status together close fully three-

⁶To the extent that marriage provides benefits to women primarily through income, but to men through other channels such as social ties and relations, it is sensible that the residual predictive power of marital status conditional on income (but with no control for social networks) is much smaller for women than for men.

quarters of the life expectancy gap for males, and just less than two-thirds of the gap for females. Urban status, which was not predictive when considered alone, becomes important after income and education are controlled for, explaining an additional 14 percentage points (76% – 62%) among men and 6 percentage points (69% – 63%) among women. Reweighting income, education, and urban residence simultaneously has the effect of scaling up the proportion of urban rich blacks and scaling down the proportion of urban poor blacks to better match the geographic distribution of income among whites. Because mortality improves, this suggests that living in an urban area is associated with a health benefit among higher-income blacks. There is no evidence, however, that urban residence confers any benefit for lower-income blacks. The results accord with Hayward et al. (1997), who found an urban mortality advantage accruing only to high-SES types.

After controlling for income and education, additionally controlling for labor force participation or white-collar status appears to have little effect for men, indicating that income and education are sufficient to capture almost all the predictive power of labor participation and white-collar status with respect to mortality.

Among women, the percentage of the gap explained decreases from a reweight on income alone to a reweight on the joint distribution of income, education, and labor force participation. This is because when the black sample is reweighted to match only the income distribution of whites, it creates a counterfactual in which black women are significantly more likely to be employed than white women. After matching on income and education, additionally matching the black sample along the dimension of labor participation results in a counterfactual sample in which fewer black women are employed compared with the income-only reweight; and because employment is associated with a mortality advantage in these data (conditional on income and education), black female life expectancy falls when employed black women are down-weighted. Thus, % explained decreases.⁷ The

⁷A similar mechanism drives the results (among women) when white-collar status is included in the reweight after conditioning on income and education.

result illustrates that decomposition procedures of this type generally impose no mechanical requirement that the percentage explained increases as more variables are added.⁸

The last two rows of Tables 2.3 and 2.4 summarize the maximum power of the control variables to account for racial differences in life expectancy at age 1. The predictive power tops off at less than 80% for men and 70% for women after jointly reweighting for income, education, marital status, and either homeownership or urban residence. Labor force participation and white-collar occupation status are left out because they have little predictive power after including income and education.

These last two rows, which match blacks to whites along four dimensions of characteristics, differ slightly from the other counterfactuals in the table because additional observations from the white sample must be dropped to maintain common support over the finer grid of characteristics. This results in a slightly different estimation sample for calculating white (but not black) life expectancy, which is an input to columns 3 and 4 of the tables. The actual values of e_1^{white} used in these two rows in each table are reported in the table notes and differ trivially from the baseline e_1^{white} in most cases.

It would be desirable to include both urban residence and homeownership simultaneously with income, education, and marriage. However, reweighting jointly along all five dimensions would result in many points at which the probability mass functions for blacks and whites lack common support. For adults alone, calculating the last row in Tables 2.3 and 2.4 without dropping observations requires observing black respondents in 2,592 cells (defined by 9 income groups $\times$ 6 education groups $\times$ 2 marital statuses $\times$ 2 urban residence statuses $\times$ 12 five-year age groups). It is possible, though, to trade off the number of variables against how finely each is coded. Therefore, I recode income into five levels (from nine) and jointly reweight on income, education, marital status, homeownership, and urban

⁸This feature is shared with other decomposition and standardization techniques. As Kitagawa (1955:1148) explained, “The difference between two crude rates is not the equivalent of a concept like total variance of a dependent variable in regression analysis, for example, which will be increasingly ‘explained’ as more independent variables are added to the regression equation.”

residence within each five-year age group. The results, not listed in a table, are consistent with Tables 2.3 and 2.4: 78% of the racial gap among males and 71% of the racial gap among females is explained.

It is straightforward to evaluate which age-specific mortality rates contribute most to the life expectancy gains resulting when the black sample is reweighted to match the white sample in socioeconomic and demographic characteristics. Applying Arriaga's (1984) technique to compare the life tables of the actual and counterfactual black samples, I find that the age-specific mortality rates mechanically responsible for the life expectancy gains under the counterfactual are concentrated primarily in the age range 40 to 69. For both men and women, more than 70% of the gains originate in this range (for a reweight on income, education, and marriage).

In Tables 2.5 and 2.6, I present the results of an exercise similar to the main analysis of Tables 2.3 and 2.4, but for which the characteristics in the white sample are made to match the characteristics in the black sample. Reweighting the white sample, in contrast with reweighting the black sample, has the feature that almost no observations need be dropped (less than 0.005%). Because white respondents outnumber black respondents roughly eight to one in the baseline survey, it is rare that a cell defined by a combination of characteristics contains a black, but not a white, respondent. Results for the white reweight are similar, but the interpretation is slightly different in that it is the relationship between SES and mortality among whites, rather than blacks, that generates the life expectancy counterfactuals.

Finally, because it may be of special interest to focus separately on adult mortality, in Tables 2.7 and 2.8 I repeat the accounting from Tables 2.3 and 2.4, but for life expectancy at age 30 rather than age 1. These results, as well, are broadly similar to the main analysis.

2.4.2 Robustness

A potential measurement concern involves the sensitivity of the results to the way in which individual variables are coded. To illustrate how the predictive power of SES can depend on the degree of flexibility in the empirical specification, I take the simple case in which income is recoded into a smaller number of wider bins. Table 2.9 presents a reweight on income only, which is coded progressively coarser moving vertically down each panel.

Each panel in Table 2.9 begins by restating the result in which income is categorized into nine groups that align roughly with deciles. The rows below it collapse income into wider bins: first quintiles, then terciles, and finally a single division at the median.⁹ For both sexes, the explanatory power of income diminishes somewhat in moving from nine groups to five groups, but moving to three groups and especially to just two groups significantly reduces the power of income to account for mortality differences. Jointly reweighting (or analogously in a regression, interacting) with a flexibly parameterized education variable could ameliorate some of the loss of predictive power because education covaries closely with income and can capture some of the income variation obscured by mismeasurement.

This characterization is relevant to the many studies that are forced by data limitations to use very coarse categorizations of SES. Table 2.9 suggests that unless the theoretically relevant variables are well measured and enter the empirical specification flexibly (regardless of whether that specification is a hazard regression or a reweight), one can easily mistake a lack of fidelity in the data for a lack of explanatory power of SES. This point about measurement error in explanatory variables is, of course, generally true, but it is particularly relevant in this context where authors are frequently constrained to enter SES as high/medium/low or high/low.

In this article, a flexible specification allows for hundreds and sometimes thousands of SES groups, each defined by a unique combination of covariates. Nonetheless, two important variables—marital status and occupation—are coded dichotomously in the main analysis.

⁹Percentiles are approximate.

This choice of coding was based on preliminary sensitivity analysis. The remainder of this section explores the consequences of allowing both variables to enter the reweight more flexibly.

To explore the robustness of the results to an alternative coding of marriage, I rerun the main analysis with marriage entered as five categories: married, divorced, widowed, separated, and never married. Given the mechanics of the decomposition technique, this requires reducing the number of income categories from nine to five to avoid dropping more than a small fraction of the white sample. For both men and women, the decompositions that included the more finely-coded marriage variable were statistically and practically indistinct from those of the main analysis. For men, marriage alone accounted for 25% of the gap (compared with 23% in Table 2.3 in the main analysis). When entered jointly with income and education, it accounted for 71% (versus 75%). For women, marriage alone accounted for 21% (versus 20% in Table 2.4) and 62% (versus 63%) when entered jointly with income and education.

A similar robustness check was performed with respect to occupation. Occupation was recoded into eight categories: managerial; professional and technical; sales; support and clerical; service; agriculture; crafts, laborers, and transportation; and missing (for those not in the labor force). Like with marriage, these results were remarkably similar to those in Tables 2.3 and 2.4. In no case for either sex was the difference between the robustness check value and the corresponding value from the main tables even marginally statistically significant.

2.5 Discussion

These results are not intended to imply that an intervention that managed to close racial gaps in education, income, and marriage rates between blacks and whites should be expected to close mortality gaps by the proportions displayed in the tables. It is doubtful that all the observed association between SES and mortality reflects a causal effect running

from SES to mortality. It is widely acknowledged, for instance, that selection into marriage is at least part of the reason for the observed association between marital status and health (see, e.g., Goldman 1993; Lillard and Panis 1996; Murray 2000). With regard to income, Case and Paxson (2011) demonstrated a channel running from health to workplace promotion and income. And since at least the Farrell and Fuchs (1982) study of education and smoking, there has been strong empirical grounding for doubts as to whether the poor health behaviors that are associated with lower education necessarily reveal anything about the causal impact of schooling on health.

To the extent that part of the correlation between SES and mortality is due to third factors and reverse causality, the decomposition results presented in this article will tend to overestimate causal effects. Thus my preferred interpretation of the results presented in the last section is as an upper bound on the causal impact of the variables considered.

Even under the most generous assumption that all the observed correlation between SES and mortality represents a causal impact of SES on health, around 20% and 30% of the racial mortality gaps among men and women, respectively, remain. The existence of a residual in related studies has been hypothesized to be attributable to causes such as differential access to quality medical care; psychosocial stress due to racial discrimination; and, most commonly, failure to capture the “right” SES variables, such as peer networks or detailed neighborhood characteristics (Williams and Collins 2001; Winkleby and Cubbin 2003). And of course, many important non-SES factors are not captured in this study (except to the extent that they covary with the variables examined). Among these are health behaviors, the most significant of which may be smoking.

This study benefits from a large data set that includes detailed measures of SES and covers mortality over a wide range of ages. Nonetheless, important limitations of the data are apparent. The analysis here has been limited to a comparison of non-Hispanic blacks and whites, both in keeping with previous literature, and due to the practical limitations imposed by small sample sizes for other race groups.

The focus on e_1 necessarily ignores the important contribution of infant mortality to life expectancy at age 0. To put this in perspective, calculations based on NCHS Vital Statistics data from the mid-1980s reveal that mortality between ages 0 and 1 accounted for around 10% of the e_0 gap between white and black males and for around 13% of the e_0 gap between white and black females in this time period. There is some external evidence that the role of SES in accounting for mortality differences between ages 0 and 1 is likely similar to my findings above age 1. For example, Finch (2003), focusing on infant mortality, found that income is the single most important predictor of infant mortality, outstripping parents' education and occupational grade in explanatory power. This suggests that with more complete data, results for a decomposition of e_0 disparities would be similar to results for e_1 .

Finally, it is important to delineate the circumstances under which a decomposition of the type found here is preferable to a hazard regression analysis. Hazards are usually the most appropriate choice when the research aim is to quantify the association between a socioeconomic variable and mortality in some age range. However, the sign, size, and significance of the coefficient on an SES variable in a regression is uninformative, in itself, as to that variable's importance in explaining group differences in mortality, which are a function both of the coefficient and of a group difference in the variable that the coefficient multiplies. Thus, the reweight technique is best viewed as an alternative to hazard regressions when the goal is decomposing group differences and individual-level data are available. It can also bypass model estimation, and it allows the decomposition of statistics other than rates and means, such as variance.

2.6 Conclusion

This article is intended to offer a straightforward answer to the question of how much of the black-white disparity in life expectancy can be accounted for by group differences in socioeconomic and demographic characteristics. The findings here that differences in such

characteristics can account for up to 80% of the gap between white and black males and up to 70% of the gap between white and black females are consistent with an extensive literature in sociology, demography, and epidemiology that finds that SES, most often operationalized as income and education, can account for a large part of the racial disparities in U.S. mortality.

Reweightings and the associated decomposition presented in this article purge mortality differences of the compositional effects of SES. Although the analysis here focuses on life expectancy (with the intention of parsimoniously summarizing the total effects of mortality differentials), the decomposition is compatible with the examination of any mortality measure, including age- and cause-specific rates. Indeed, further decomposing these results into contributions of specific causes of death is an interesting avenue of future research.

Table 2.1: Summary statistics at baseline

	White		Black	
	Male	Female	Male	Female
Panel A: All Respondents				
Persons	372,639	400,209	42,880	53,356
Deaths	40,731	35,470	4,610	4,307
Family income (1990 \$)				
<10,000	0.08	0.11	0.26	0.32
10,000-19,999	0.15	0.17	0.25	0.26
20,000-29,999	0.21	0.21	0.21	0.19
30,000-39,999	0.17	0.16	0.12	0.10
40,000-74,999	0.30	0.27	0.14	0.12
≥75,000	0.08	0.07	0.02	0.02
Household size	3.53	3.39	4.16	4.14
Own home	0.75	0.73	0.50	0.47
Urban	0.64	0.66	0.83	0.84
Age	32.58	34.89	27.4	29.7
Panel B: Respondents Aged 25 and Older at Baseline				
Persons	222,590	252,307	20,255	28,607
Education (years)				
Elementary (0-4)	0.02	0.01	0.09	0.05
Middle school (5-8)	0.11	0.11	0.16	0.16
Some high school (9-11)	0.11	0.13	0.19	0.21
High school (12)	0.35	0.44	0.32	0.34
Some college (13-15)	0.17	0.16	0.14	0.14
College and graduate school (16+)	0.24	0.16	0.10	0.09
In labor force	0.80	0.52	0.75	0.58
Employed	0.90	0.89	0.83	0.82
White-collar job	0.49	0.72	0.26	0.49
Currently married	0.78	0.69	0.59	0.43
Age	46.25	48.02	44.66	45.26

Notes: Income deflated to 1990 dollars. Except for the variables persons, deaths, household size, and age, the numbers reported are fractions. Employment and white-collar status are calculated conditional on labor force participation.

Table 2.2: Life expectancy comparability between the NLMS and NCHS

Life Expectancy at Age 1 (e_1)	White		Black	
	Male	Female	Male	Female
NLMS	71.3	78.1	66.6	74.3
NCHS, 1987	71.9	78.5	65.5	73.8
NCHS, 1988	72.0	78.5	65.2	73.6

Sources: NLMS estimates from author's calculations. NCHS estimates are from vital statistics (NCHS 1990, 1991).

Table 2.3: Decomposition of e_1 differences by SES and demographics: Results for males, reweighting the black sample

	(1) e_1	(2) $SE(e_1)$	(3) $e_1^{white} - e_1$	(4) % Explained	(5) Reweight Numerator
White	71.2	0.07	0		
Black	66.64	0.28	4.56		
Black Reweight on					
Income (inc)	69.02	0.38	2.17	52	f(inc a, R = w)
Education (ed)	67.83	0.29	3.37	26	f(ed a, R = w)
Marital status (ms)	67.71	0.29	3.49	23	f(ms a, R = w)
Own home (own)	67.37	0.35	3.83	16	f(own a, R = w)
In labor force (lf)	67.59	0.29	3.61	21	f(lf a, R = w)
White-collar job (wc)	67.89	0.29	3.31	27	f(wc a, R = w)
Urban residence (urb)	66.76	0.28	4.44	3	f(urb a, R = w)
Inc, ed	69.45	0.39	1.75	62	f(inc, ed a, R = w)
Ed, ms	68.79	0.29	2.41	47	f(ed, ms a, R = w)
Inc, ms	69.67	0.39	1.53	66	f(inc, ms a, R = w)
Inc, ed, ms	70.06	0.39	1.14	75	f(inc, ed, ms a, R = w)
Inc, ed, own	69.82	0.44	1.38	70	f(inc, ed, own a, R = w)
Inc, ed, lf	69.54	0.39	1.65	64	f(inc, ed, lf a, R = w)
Inc, ed, wc	69.49	0.39	1.71	62	f(inc, ed, wc a, R = w)
Inc, ed, urb	70.09	0.40	1.11	76	f(inc, ed, urb a, R = w)
Inc, ed, ms, own	70.18	0.44	1.10	76	f(inc, ed, ms, own a, R = w)
Inc, ed, ms, urb	70.19	0.40	1.04	77	f(inc, ed, ms, urb a, R = w)

Notes: Column 1 gives actual and counterfactual values for life expectancy at age 1. Standard errors in column 2 are calculated following Chiang (1968). Column 3 displays the difference between column 1 and the value of e_1 in the white subsample (e_1^{white}). In all but the last two rows, the value of e_1^{white} used throughout the table to calculate columns 3 and 4 is the value reported in the first row of column 1. The values of e_1^{white} used for calculations in the last two rows are 71.28 and 71.23, respectively. See the text for further details.

Table 2.4: Decomposition of e_1 differences by SES and demographics: Results for females, reweighting the black sample

	(1) e_1	(2) $SE(e_1)$	(3) $e_1^{white} - e_1$	(4) % Explained	(5) Reweight Numerator
White	77.91	0.07	0		
Black	74.34	0.23	3.57		
Black Reweight on					
Income (inc)	76.46	0.27	1.44	59	$f(\text{inc} \mid a, R = w)$
Education (ed)	75.37	0.23	2.54	29	$f(\text{ed} \mid a, R = w)$
Marital status (ms)	75.07	0.23	2.84	20	$f(\text{ms} \mid a, R = w)$
Own home (own)	75.15	0.23	2.76	23	$f(\text{own} \mid a, R = w)$
In labor force (lf)	74.09	0.23	3.82	7	$f(\text{lf} \mid a, R = w)$
White-collar job (wc)	74.44	0.23	3.47	3	$f(\text{wc} \mid a, R = w)$
Urban residence (urb)	74.42	0.23	3.49	2	$f(\text{urb} \mid a, R = w)$
Inc, ed	76.59	0.27	1.32	63	$f(\text{inc, ed} \mid a, R = w)$
Ed, ms	75.83	0.23	2.08	42	$f(\text{ed, ms} \mid a, R = w)$
Inc, ms	76.66	0.27	1.25	65	$f(\text{inc, ms} \mid a, R = w)$
Inc, ed, ms	76.58	0.26	1.33	63	$f(\text{inc, ed, ms} \mid a, R = w)$
Inc, ed, own	76.73	0.25	1.18	67	$f(\text{inc, ed, own} \mid a, R = w)$
Inc, ed, lf	76.11	0.27	1.80	49	$f(\text{inc, ed, lf} \mid a, R = w)$
Inc, ed, wc	76.29	0.27	1.61	55	$f(\text{inc, ed, wc} \mid a, R = w)$
Inc, ed, urb	76.79	0.28	1.12	69	$f(\text{inc, ed, urb} \mid a, R = w)$
Inc, ed, ms, own	76.77	0.25	1.25	66	$f(\text{inc, ed, ms, own} \mid a, R = w)$
Inc, ed, ms, urb	76.69	0.27	1.24	65	$f(\text{inc, ed, ms, urb} \mid a, R = w)$

Notes: Column 1 gives actual and counterfactual values for life expectancy at age 1. Standard errors in column 2 are calculated following Chiang (1968). Column 3 displays the difference between column 1 and the value of e_1 in the white subsample (e_1^{white}). In all but the last two rows, the value of e_1^{white} used throughout the table to calculate columns 3 and 4 is the value reported in the first row of column 1. The values of e_1^{white} used for calculations in the last two rows are 78.02 and 77.93, respectively. See the text for further details.

Table 2.5: Decomposition of e_1 differences by SES and demographics: Results for males, reweighting white sample

	(1) e_1	(2) $SE(e_1)$	(3) $e_1 - e_1^{black}$	(4) % Explained	(5) Reweight Numerator
Black	66.64	0.28	0		
White	71.31	0.07	4.67		
White Reweight on					
Income (inc)	69.17	0.09	2.53	46	f(inc a, R = b)
Education (ed)	69.86	0.07	3.22	31	f(ed a, R = b)
Marital status (ms)	70.35	0.07	3.71	21	f(ms a, R = b)
Own home (own)	70.49	0.08	3.85	18	f(own a, R = b)
In labor force (lf)	70.36	0.07	3.72	20	f(lf a, R = b)
White-collar job (wc)	70.02	0.07	3.38	28	f(wc a, R = b)
Urban residence (urb)	71.30	0.07	4.66	0	f(urb a, R = b)
Inc, ed	68.69	0.09	2.05	56	f(inc, ed a, R = b)
Ed, ms	69.01	0.07	2.37	49	f(ed, ms a, R = b)
Inc, ms	68.53	0.09	1.89	60	f(inc, ms a, R = b)
Inc, ed, ms	68.07	0.09	1.43	69	f(inc, ed, ms a, R = b)
Inc, ed, own	68.29	0.08	1.65	65	f(inc, ed, own a, R = b)
Inc, ed, lf	68.60	0.09	1.96	58	f(inc, ed, lf a, R = b)
Inc, ed, wc	68.60	0.09	1.96	58	f(inc, ed, wc a, R = b)
Inc, ed, urb	68.39	0.08	1.75	63	f(inc, ed, urb a, R = b)
Inc, ed, ms, own	67.89	0.08	1.25	73	f(inc, ed, ms, own a, R = b)
Inc, ed, ms, urb	67.91	0.08	1.27	73	f(inc, ed, ms, urb a, R = b)

Notes: Standard errors in column 2 are calculated following Chiang (1968). Throughout the table, the value of e_1^{black} used to calculate columns 3 and 4 is the value reported in the first row of column 1.

Table 2.6: Decomposition of e_1 differences by SES and demographics: Results for females, reweighting white sample

	(1) e_1	(2) $SE(e_1)$	(3) $e_1 - e_1^{black}$	(4) % Explained	(5) Reweight Numerator
Black	74.34	0.23	0		
White	78.06	0.07	3.72		
White Reweight on					
Income (inc)	76.39	0.07	2.05	45	f(inc a, R = b)
Education (ed)	77.11	0.07	2.77	26	f(ed a, R = b)
Marital status (ms)	77.35	0.07	3.01	19	f(ms a, R = b)
Own home (own)	77.36	0.07	3.02	19	f(own a, R = b)
In labor force (lf)	78.13	0.07	3.79	-2	f(lf a, R = b)
White-collar job (wc)	77.99	0.07	3.64	2	f(wc a, R = b)
Urban residence (urb)	77.99	0.07	3.65	2	f(urb a, R = b)
Inc, ed	76.01	0.07	1.67	55	f(inc, ed a, R = b)
Ed, ms	76.45	0.07	2.11	43	f(ed, ms a, R = b)
Inc, ms	76.08	0.07	1.74	53	f(inc, ms a, R = b)
Inc, ed, ms	75.71	0.07	1.37	63	f(inc, ed, ms a, R = b)
Inc, ed, own	75.67	0.07	1.33	64	f(inc, ed, own a, R = b)
Inc, ed, lf	76.33	0.07	1.99	47	f(inc, ed, lf a, R = b)
Inc, ed, wc	76.27	0.07	1.93	48	f(inc, ed, wc a, R = b)
Inc, ed, urb	75.69	0.07	1.34	64	f(inc, ed, urb a, R = b)
Inc, ed, ms, own	75.50	0.07	1.16	69	f(inc, ed, ms, own a, R = b)
Inc, ed, ms, urb	75.47	0.07	1.13	70	f(inc, ed, ms, urb a, R = b)

Notes: Standard errors in column 2 are calculated following Chiang (1968). Throughout the table, the value of e_1^{black} used to calculate columns 3 and 4 is the value reported in the first row of column 1.

Table 2.7: Decomposition of e_{30} differences by SES and demographics: Results for males, reweighting the black sample

	(1) e_{30}	(2) $SE(e_{30})$	(3) $e_{30}^{white} - e_{30}$	(4) % Explained	(5) Reweight Numerator
White	43.17	0.05	0		
Black	39	0.2	4.17		
Black Reweight on					
Income (inc)	41.26	0.21	1.91	54	f(inc a, R = w)
Education (ed)	40.21	0.21	2.96	29	f(ed a, R = w)
Marital status (ms)	40.09	0.20	3.09	26	f(ms a, R = w)
Own home (own)	39.68	0.20	3.49	16	f(own a, R = w)
In labor force (lf)	39.97	0.20	3.20	23	f(lf a, R = w)
White-collar job (wc)	40.27	0.21	2.90	30	f(wc a, R = w)
Urban residence (urb)	39.06	0.20	4.11	1	f(urb a, R = w)
Inc, ed	41.69	0.22	1.48	64	f(inc, ed a, R = w)
Ed, ms	41.18	0.21	1.99	52	f(ed, ms a, R = w)
Inc, ms	41.90	0.21	1.27	70	f(inc, ms a, R = w)
Inc, ed, ms	42.29	0.22	0.88	79	f(inc, ed, ms a, R = w)
Inc, ed, own	42.08	0.22	1.09	74	f(inc, ed, own a, R = w)
Inc, ed, lf	41.77	0.22	1.40	66	f(inc, ed, lf a, R = w)
Inc, ed, wc	41.71	0.21	1.46	65	f(inc, ed, wc a, R = w)
Inc, ed, urb	42.35	0.22	0.82	80	f(inc, ed, urb a, R = w)
Inc, ed, ms, own	42.42	0.21	0.84	84	f(inc, ed, ms, own a, R = w)
Inc, ed, ms, urb	42.43	0.21	0.78	81	f(inc, ed, ms, urb a, R = w)

Notes: Standard errors in column 2 are calculated following Chiang (1968). In all but the last two rows, the value of e_{30}^{white} used throughout the table to calculate columns 3 and 4 is the value reported in the first row of column 1. The values of e_{30}^{white} used for calculations in the last two rows are 43.26 and 43.21, respectively.

Table 2.8: Decomposition of e_{30} differences by SES and demographics: Results for females, reweighting the black sample

	(1) e_{30}	(2) $SE(e_{30})$	(3) $e_{30}^{white} - e_{30}$	(4) % Explained	(5) Reweight Numerator
White	49.45	0.06	0		
Black	46.13	0.21	3.32		
Black Reweight on					
Income (inc)	48.18	0.22	1.27	62	f(inc a, R = w)
Education (ed)	47.16	0.21	2.29	31	f(ed a, R = w)
Marital status (ms)	46.85	0.21	2.6	22	f(ms a, R = w)
Own home (own)	46.83	0.21	2.62	21	f(own a, R = w)
In labor force (lf)	45.87	0.21	3.58	-8	f(lf a, R = w)
White-collar job (wc)	46.22	0.21	3.23	3	f(wc a, R = w)
Urban residence (urb)	46.18	0.21	3.27	2	f(urb a, R = w)
Inc, ed	48.30	0.22	1.15	65	f(inc, ed a, R = w)
Ed, ms	47.62	0.22	1.83	45	f(ed, ms a, R = w)
Inc, ms	48.37	0.22	1.08	68	f(inc, ms a, R = w)
Inc, ed, ms	48.28	0.22	1.17	65	f(inc, ed, ms a, R = w)
Inc, ed, own	48.37	0.22	1.08	68	f(inc, ed, own a, R = w)
Inc, ed, lf	47.83	0.22	1.62	51	f(inc, ed, lf a, R = w)
Inc, ed, wc	48.00	0.22	1.45	56	f(inc, ed, wc a, R = w)
Inc, ed, urb	48.54	0.22	0.91	73	f(inc, ed, urb a, R = w)
Inc, ed, ms, own	48.41	0.21	1.15	66	f(inc, ed, ms, own a, R = w)
Inc, ed, ms, urb	48.44	0.22	1.03	69	f(inc, ed, ms, urb a, R = w)

Notes: Standard errors in column 2 are calculated following Chiang (1968). In all but the last two rows, the value of e_{30}^{white} used throughout the table to calculate columns 3 and 4 is the value reported in the first row of column 1. The values of e_{30}^{white} used for calculations in the last two rows are 49.56 and 49.47, respectively.

Table 2.9: Sensitivity analysis: Alternative categorizations of income

	e_1	$SE(e_1)$	$e_1^{white} - e_1$	% Explained
Panel A: Males				
White	71.20	0.07		
Black	66.64	0.28	4.56	
Black reweight on				
Income (9 groups)	69.02	0.38	2.17	52
Income (5 groups)	68.98	0.34	2.22	51
Income (3 groups)	68.76	0.32	2.44	46
Income (2 groups)	68.24	0.31	2.96	35
Panel B: Females				
White	77.91	0.07		
Black	74.34	0.23	3.57	
Black reweight on				
Income (9 groups)	76.46	0.27	1.44	59
Income (5 groups)	76.21	0.24	1.70	52
Income (3 groups)	76.07	0.24	1.83	49
Income (2 groups)	75.62	0.24	2.28	36

Notes: Standard errors are calculated following Chiang (1968).

Chapter 3

The Impact of Education on Fertility: Quasi-Experimental Evidence from the UK

3.1 Introduction

In many OECD countries, completed fertility is low—both by historical standards and relative to the replacement levels needed to maintain stable populations.¹ Within these countries, one of the most robust predictors of individual-level fertility is a woman’s educational attainment. If the relationship between education and fertility is causal, then the well-known benefits of further increases in educational attainment could be accompanied by real costs in terms of lower fertility rates. Such a relationship would be particularly troubling for areas such as Europe where an already aging population is expected to have serious impacts on generous pay-as-you-go pension systems and public health systems.

Aside from completed fertility, education may affect fertility timing. One important dimension is teen fertility, an outcome considered bad for both teen mothers Geronimus and Korenman (1992); Goodman et al. (2004); Chevalier et al. (2003); Ashcraft and Lang (2006), and their children Royer (2004); Painter and Levine (2004). Given the strong inverse correlation between educational attainment and teen fertility, it is not surprising

¹We thank Julie Jefferies and Louise O’Leary at the Office of National Statistics for the considerable time and effort that they and their team have devoted to this project. It would not have been possible without the live births dataset that they supplied. We also thank Mary Grinsted for generously supplying the abortions data.

therefore that education is regarded as a policy lever for reducing teen fertility.²

While partial correlations between education and various fertility outcomes raise the possibility that education affects fertility, they do not necessarily imply a causal connection. The omitted variables problems that are pervasive throughout empirical studies are likely prominent here. For example, a female from a disadvantaged family is at higher risk of a teenage pregnancy Geronimus and Korenman (1992), as well as at a higher risk of early exit from school.

Economic theory does not provide clear guidance as to direction of the effect of education on fertility. Theory emphasizes mechanisms pushing in both directions. On the one hand, more-educated women will have better labor-market opportunities Becker (1960); Mincer (1963); Willis (1973). If participation in the labor market precludes women from bearing and raising children, this implies that more-educated women will have lower completed fertility and likely delayed fertility. On the other hand, to the extent that children represent a normal good, wealthier women will choose to have more of them. More-educated women may be wealthier because education has a direct effect on their own earnings and/or because it has an indirect effect on their husband's earnings via assortative matching in the marriage market Becker (1981).

In light of the strong correlation between education and other characteristics influencing fertility, it is not surprising that the empirical evidence on the education-fertility relationship is inconclusive. Black et al. (2008) uncover large teen fertility effects exploiting compulsory schooling reforms in the US and Norway. In contrast, McCrary (2011) find no effects when they utilize variation in educational attainment due to school starting rules in the US. Currie and Moretti (2003) estimate significant fertility responses to college openings but have crude measures of fertility.³ However, this literature suffers from several

²See for instance, the website for the National Campaign to Prevent Teen Pregnancy (<http://www.thenationalcampaign.org>).

³Within the economics literature, empirical work on the fertility effects of education is more complete for developing countries, where recent large national pushes to increase education can be used to identify the fertility effects of education. Examples include Breierova and Duflo (2004), Osili and Long (2008), and

weaknesses. For one, these three papers are limited in the types of fertility outcomes they examine—none evaluates effects on completed fertility. Also, much of the literature relies on difference-in-difference analyses, which are subject to endogeneity concerns if trends are not adequately controlled for.⁴ A more recent literature attempts to address this empirical challenge by employing regression discontinuity designs, which allow for a direct visual inspection of the adequacy of trend controls. For example, McCrary (2011) exploit a discontinuity in exact date of birth—though they are unable to examine the impact of completed education and are limited to measuring fertility over a narrow age range.

In this paper, we estimate the causal effect of education on both completed fertility and teen fertility. Our analysis can shed valuable new light on this empirical question for several reasons. First, we exploit sharp and plausibly exogenous variation in education. Specifically, we use a 1972 UK education reform that raised the minimum school-leaving age in England and Wales from 15 to 16. The change affected all women born in September 1957 or later. As such, we can rely on regression discontinuity (RD) comparisons of women born very close together (e.g., September 1957 versus August 1957) who might be otherwise comparable aside from their education. In other quasi-experiments, RD estimators are not often feasible, either because of the resolution of the data or because the implementation details don't permit it (e.g., in the United States, Lleras-Muney 2005).

Second, the reform impacted a large fraction of the population: roughly one quarter of the women born in September 1957 received an additional year of schooling as a result of the reform. This allows us to estimate the effects of this reform very precisely. It also means that our local average treatment effects are estimated for a large and policy-relevant group, somewhat in contrast to settings such as the US, where compulsory schooling laws have been binding for only a few percent of young people Goldin and Katz (2008).

Duflo et al. (2011). These still tend to focus on teen outcomes, in part because the affected cohorts are too young to have completed their fertility.

⁴See for instance, Mazumder (2008) who shows that the use of state-level compulsory schooling changes as instruments for education is not robust to the inclusion of state-specific time trends.

Third, the reform was recent and thus the institutional features, such as access to legal abortion and modern contraception, were quite similar to today, allowing us to examine these effects in a modern institutional context. Yet enough time has elapsed since the reform for us to explore the completed fertility effects.

Finally, in comparison to the prior literature, we can better address the effects of education on conceptions, as opposed to solely on births, by examining impacts on abortions. We combine information on educational attainment from the UK Labor Force Survey with administrative records on the universe of live births and abortions in England and Wales.

Our analysis reveals three main findings. First, an additional year of schooling had a significant impact on teen fertility, reducing it by around 20% at ages 16 and 17. Second, the additional schooling had at most a negligible effect on teen abortions. This implies that the teen fertility effects reflect a reduction in conceptions. Third, we cannot reject that an additional year of schooling had no impact on post-teen fertility and no impact on completed family size.⁵

We discuss several explanations for why this educational reform could have reduced teen fertility. Crucially, we show that the timing of these teen fertility reductions is consistent with education reducing fertility by reducing the probability that girls conceive while they are in school. We also discuss why this additional education appears not to have changed total fertility. From a theoretical perspective, the most obvious explanation is that the income and substitution effects of the additional earnings enjoyed by more-educated women roughly cancelled out. While Willis (1973) suggests that the substitution effect is more likely to dominate, that analysis ignores the effects of women's education on husband's income, a mechanism discussed by Becker (1981) and documented by Lefgren and McIntyre (2006). To explore this issue, we present evidence of the reform's effect on women's marriage outcomes, including husband's education.

⁵While the teen fertility reductions are large as a percentage of teen births, teen births are such a small part of overall births that complete pass through of the teen effects into completed fertility would fall well within the tight confidence intervals of the completed fertility effects.

With the important caveat that we study only one margin of education, the effects of which may be different from those at other margins (e.g., years in college), we conclude that there is no evidence that a trend to increased education in the OECD need imply reductions in future levels of completed fertility. Nevertheless, our study does suggest that schooling can be an effective tool in the fight against teen fertility in developed countries.⁶ This is important both because teen fertility reduction is often a difficult policy goal to achieve and because, as discussed below, policy-makers in the UK, US, and other countries are currently considering measures that would compel girls to stay in school for longer.

3.2 Compulsory Schooling Laws and Fertility Trends

3.2.1 Compulsory Schooling Laws

Compulsory schooling laws in the UK are set at the national level.⁷ They include a maximum age by which children must start school and a minimum age at which children can leave school. The maximum starting age is five and has been so for over a century. In contrast, the minimum leaving age was changed twice in the postwar period. These changes, also known as ROSLAs (Raising of the School Leaving Age), were implemented in 1947 and 1972. Further increases are slotted for the near future.⁸

The 1947 ROSLA increased the minimum leaving age from 14 to 15. Several papers have used this reform to study the effects of education on various outcomes (e.g., Harmon and Walker 1995; Oreopoulos 2006; Devereux and Hart 2010; Clark and Royer 2010). The more recent 1972 ROSLA increased the minimum leaving age from 15 to 16. This second ROSLA has received less attention in the literature; Harmon and Walker (1995) Clark and Royer (2010), and Machin et al. (2011) use it to study the impact of education on wages,

⁶Teen fertility reduction is often cited as an important motivation for increasing educational attainment among women in developing countries.

⁷Much of this discussion is adapted from Clark and Royer (2010).

⁸The Education and Skills Act 2008 stipulates that individuals remain in schooling or training until the age of 17 starting in 2013 and until the age of 18 beginning in 2015. See website: <http://www.politics.co.uk/reference/education-leaving-age>.

health, and crime, respectively. We focus exclusively on the 1972 ROSLA as because fertility data are not available for the women affected by the 1947 ROSLA.

The 1972 ROSLA applied to students in England, Wales and Scotland. We focus on the effects for women born in England and Wales because we do not have fertility data for Scotland. The government announced the change in March 1972 and on 1 September 1972, the change became effective. Students who turned 15 before September 1972 (born August 1957 or earlier) were subject to the old law; students who turned 15 after this date (born September 1957 or later) were subject to the new law. Though the change was officially announced in 1972, plans and preparations had begun as early as 1964.

A detail particularly relevant for our study is the treatment of pregnant females within the schooling system. Both before and after the ROSLA, girls who became pregnant during the period of compulsory schooling were required to continue with education through their pregnancy. They attended school until the later months of their pregnancy, after which they received an in-home, state-provided tutor. Thus a teen pregnancy did not result in a waiver from the compulsory schooling requirement.

In Figure 3.1, we examine visually the impact of the 1972 ROSLA on age at school exit. Specifically, each scatter point in Figure 3.1 represents the average outcome for a cohort of women who share the same month and year of birth. The three sets of points trace the fraction of women who left school by age 15, by age 16, and by age 17.

In this figure, we observe that the compulsory schooling law was binding for a significant share of individuals. Almost one third of women born before September 1957 left school at age 15.⁹ The reform had an immediate impact on age at school exit as there is a clear drop in the fraction leaving before age 16 (i.e., before grade 10 in a US context) starting with the September 1957 cohort. This gap is on the order of 25 percentage points. Effects for men (not pictured) are similar.

While Figure 3.1 suggests that most students complied with the new law, there is a non-

⁹The data underlying this figure—the UK Labor Force Survey—are discussed in more detail below.

negligible share of individuals who report not complying (18% for most cohorts). Analyses of Labor Force Study data suggest that some of the apparent non-compliance likely reflects errors in self-reported school leaving age. Because the LFS is a panel, respondents are asked about school leaving multiple times, and over 10% of respondents who report leaving school at age 15 also report leaving school at 16 in a different survey wave. There is also an apparent seasonal pattern of non-compliance among cohorts born in June, July, and especially August. This is because school was not in session when these women turned 16. Hence they report leaving school at 15, even though they are likely to have completed grade 10, in compliance with the law.

Figure 3.1 shows that the reform only induced those at risk of dropping out at age 15 to stay in school an additional year. We do not find any effects of the reform on the probability of attending schooling beyond the age of 16. In the analyses below, we examine these effects within a regression framework.

3.2.2 Fertility Trends

Fertility declined dramatically between the mid-1960s and the mid-1970s in the UK. Figure 3.2 displays the trend, along with trends in the US, France and Germany for comparison. In the UK, by 1975, total fertility had fallen to modern, below-replacement levels.

The post-1972 ROSLA period in the UK, marked by the vertical line, is a period of low and stable fertility (1972-2002). Modern fertility rates in the UK are quite comparable to those of the US and France. Given the similarity of fertility today and at the time of the enactment of the 1972 ROSLA, the 1972 ROSLA is almost ideal for studying the education-fertility relationship: early enough for the affected women to have now completed their fertility, but late enough for them to be informative about childbearing in a modern setting. Moreover, the cohorts just affected by the reform have shared the same access to modern contraception and abortion services throughout their fecund lives as women

today.¹⁰

Aside from total fertility, we examine teenage births, often an explicit target of fertility policies. As seen in Figure 3.3, teen fertility also declined through the 1960s and 1970s, although in the UK it fell slightly slower and slightly later than total fertility. The timing of the fertility decline is consistent with a possible effect of the 1972 ROSLA. However, all countries depicted in the graph exhibit a similarly-timed decline, raising the possibility that wider trends could be confounded with a ROSLA program effect. This highlights the importance of using a research design that can credibly control for these cohort and time trends. Teen fertility in the UK stabilized at a level higher than those seen elsewhere in Western Europe but below that seen in the US.¹¹ As of 2009, about one in eight girls in the UK bore a child before age 20.

Given the oft-mentioned detrimental consequences of teen pregnancy and given the problems associated with an aging population, governments are in search of potential fertility policy levers. The UK government has consistently sought to lower teen fertility and in 1998 launched a £260m (roughly 425m in 2010 US dollars) campaign to halve teen fertility by 2010. Similar efforts and funding commitments have been made in the US. In contrast, governments often view completed fertility as being too low since below replacement rates imply slower economic growth and increased burdens on entitlement programs. In response many European governments have introduced policies designed to encourage larger families. These include reforms to maternity pay and paternity leave, more flexible hours, and a family tax break (See, for example, Kalwij (2010) for a discussion of these types of pro-natal policies in Western European countries.) With these divergent policy goals in mind, an interesting question is whether schooling can reduce teen fertility without lowering total fertility.

¹⁰The 1967 Abortion Act legalized abortion in the UK. This Act covered abortions for teenage girls and did not require parental consent or notification.

¹¹The Healthy People 2010 and Healthy People 2020 initiatives, within the US Department of Health and Human Services, both set targets for reducing the pregnancy rate among adolescent females aged 15 to 17 years.

3.3 Data

Our study uses several datasets. We describe each of these in turn. The common variables linking our datasets are the month and year of birth of women whose education, fertility, and partnership outcomes we examine. Our analysis is at the level of month-of-birth cohorts, as we describe later. A summary of the data sources and the relevant variables we extract from each is provided in Table 3.1.

UK Labor Force Survey. Information on educational attainment comes from the 1975-2002 UK Labor Force Survey (LFS). The LFS samples a large number of households to provide information on labor force participation, employment, and training. Around 60,000 households are interviewed in each wave. Between 1975 and 1983 the LFS was conducted biannually. Between 1983 and 1992 the LFS was an annual survey of a similar number of households. Since 1992, the survey has been a rotating quarterly panel, with the same households interviewed for five consecutive quarters. In the analysis, we include multiple observations per person when available, and cluster all results at the cohort level. Respondents' year and month of birth are available in public release LFS datasets up to 2002.

Since its existence, LFS respondents have been asked about the age at which they first left full-time continuous education. Respondents were also asked about the highest academic qualifications they held, though the exact form of these questions differs across LFS survey years. Our pooled sample of 1975-2002 LFS gives us a large sample of over 300,000 respondents born between 1954 and 1962. These data generate Figure 3.1, which measures the impact of ROSLA on educational attainment.

Live Birth Records. Live birth data were supplied by the UK Office for National Statistics (ONS). The records cover every birth between 1970 and 2008 in England and Wales to mothers born in England, Wales, or Scotland—roughly 600,000 to 900,000 births per year.¹² Since teen fertility is a relatively rare outcome, it is crucial to have a census of

¹²The birth records from England and Wales do not distinguish between mothers born in England and

births. The births data are birth counts by mother’s month and year of birth and child’s month and year of birth but do not include parity information. Given that we rely on across-cohort variation in education, under most scenarios, it is not problematic that we do not have individual-level birth data. Specifically, our regression framework, described later, does not require the inclusion of covariates aside from controls for cohort trends. Any covariates, in principle, should be similar for females born immediately prior to September 1957 and females born immediately following September 1957.

The structure of our data implies that we can recover cohort-specific fertility rates but not individual-level fertility rates. For instance, our data cannot distinguish between a scenario in which one woman bears her first child at age 20 and another bears her first child at age 24 and a scenario in which one woman bears her first child at age 20 and then her second at age 24. In practice, since we find that the reform only impacts fertility among girls aged 16 and 17, who are very likely to have been first-time mothers, the missing parity information is of little consequence. Moreover, assuming that education affects fertility behavior mainly in one direction (i.e., either increasing it or decreasing it), the inability to distinguish parity will not cause us to mischaracterize the results.¹³

For each cohort of women, we construct age-specific and total fertility rates according to standard practice (Preston et al. 2001). The numerator in the age-specific rate for birth cohort c is the number of births to women of cohort c over a particular age range (e.g., [16, 17)). The denominator is the number of women alive in that age range. Unfortunately, it is not possible to know exactly how many women are alive at each point in time. Thus, we generate the denominator by multiplying the size of the female cohort at birth (obtained from the Registrar’s Annual Reports) with a succession of age- and calendar year-specific

Wales and mothers born in Scotland. This poses minimal problems for our study, since only a small fraction of the women living in England and Wales during this time were emigrants from Scotland ($< 1\%$).

¹³For example, counter to this assumption, suppose that the reform influences one woman to increase her fertility preferences from 2 to 3 children and influences another to reduce her preferences from 1 child to being childless. In this case, the total number of children born is unchanged but this overall effect masks important heterogeneity.

survivor rates (obtained from the Human Mortality Database). We validate these denominator estimates by comparing our predictions for each cohort size in 1991 and 2001 with cohort size as measured by the 1991 and 2001 Censuses. Our estimates and these numbers line up very closely (details available from the authors). The total fertility rate (TFR) aggregates the age specific rates. Cohort TFR reflects the average number of children that women in the cohort bear over their lifetime, conditional on surviving through their fecund years.

Abortion Records. Detailed abortion records come from the UK Office of National Statistics. Like the live birth records, these data are counts by women’s month and year of birth. Unlike the live birth records, these are limited to a narrower range of cohorts born around the reform threshold. They also cover only abortions performed in the teen years. Nevertheless, these data allow us to construct teen (16-19) abortion rates for cohorts of women born within a 36-month window of the ROSLA threshold.

UK Census Extracts. Finally, we commissioned multiple extracts of UK Census data. An extract of the 1991 Census provides information on the marital status of all Census respondents. We also commissioned an extract of the 1991 Census that provides information on the age difference between partners in all marriages in which the partners live in the same household. The first extract allows us to determine whether the ROSLA affected the probability of marriage. The second extract allows us to examine the assortative mating effects of the ROSLA.

National Child Development Study. In future versions of the paper, we will use data from the National Child Development Study (NCDS), a panel that attempted to track 100% of babies born in Great Britain in a particular week of March 1958. This cohort was one of the first to be subject to the 1972 ROSLA. Information collected at ages 7, 11, 16, 23, 33, and 41 will offer insight into cohort members’ school activities, partnerships, and labor force participation during the extra year of compulsory schooling and beyond.

3.4 Empirical Strategy

To exploit the natural experiment represented by the 1972 ROSLA, we estimate versions of the following equations:

$$E_{ic} = \gamma_0 + \gamma_1 D_c + f(c) + u_{ic} \quad (3.1)$$

$$Y_{ic} = \beta_0 + \beta_1 D_c + g(c) + v_{ic} \quad (3.2)$$

where E_{ic} is an indicator for whether individual i in cohort c experienced a year of full-time education at age 15, Y_{ic} is a fertility outcome, D_c is an indicator for whether the cohort is affected by the 1972 ROSLA, and $f(c)$ and $g(c)$ are smooth functions of the birth cohorts (low-order polynomials). Cohorts are groups of women born in the same calendar month of the same year. The error terms, u_{ic} and v_{ic} capture the unobservable determinants of E_{ic} and Y_{ic} respectively. The inclusion of $f(c)$ and $g(c)$ will capture underlying time or cohort trends in education and fertility that evolve smoothly.

Because births and abortions data contain counts by month and year of birth, we aggregate to get month-of-birth means in all datasets. We then estimate cohort versions of these equations, weighted by cohort size. In practice, we normalize the cohort variable c so that it equals zero for women born in September 1957, one for women born in October 1957, and so on. This normalization implies that coefficient estimates of γ_1 and β_1 in the regression tables can be directly interpreted as estimates of the ROSLA impact.

The parameters of interest are γ_1 and β_1 . They represent, respectively, the ROSLA effects on the probability that a girl experiences a year of education between age 15 and 16 and on the various fertility outcomes. These parameters can be thought of as the “jump” in the dependent variable at the program threshold. This is the sense in which our research design exploits the comparison of cohorts “just unaffected” and “just affected” by the reform. The identifying assumption underpinning this strategy is that there would have

been no such jump in the absence of the 1972 ROSLA. The data sources for education, births, and abortions are separate, so we separately estimate reduced form effects of ROSLA on these outcomes. The fertility effects can be scaled by the first-stage schooling estimates to provide an instrumental variables estimate of the effect of an additional year of education on various fertility outcomes. The ratio β_1/γ_1 is the standard Wald IV estimate in a two-sample instrumental variables strategy (Angrist and Krueger 1992).

To estimate these equations, we must select the specification of the cohort trend functions g and f and the cohort window over which to estimate them. For most outcomes, we specify a linear cohort trend interacted with a ROSLA dummy. The interaction allows trends to evolve differently in the pre and post ROSLA environments. We choose the cohort window on either side of the reform according to a cross-validation procedure outlined in Lee and Lemieux (2010). This chooses the window that provides the best trade-off between precision (which results from wide windows) and bias (which is minimized with narrow windows). Since our estimates may be sensitive to these choices, we conduct several robustness checks. First, to provide visual evidence of any specification errors, we graph the relationship between the running variable (birth cohort) and each outcome. Second, we experiment with both alternative specifications of the cohort trends (e.g., quadratic, cubic, etc) and with both narrower and wider cohort windows. These specification checks are reported in appendix tables.

An identifying assumption under which our IV estimates are valid measures of the effect of a year of additional schooling at age 15 is that the 1972 ROSLA does not affect fertility through channels other than its effect on the probability of experiencing a year of full-time education at age 15. This exclusion restriction would be violated if the law had impacts on attainment further up the education distribution. For example, by compelling girls who would previously have left school at age 15 to stay in school until age 16, it may encourage some to go further and stay until 17 or 18. Alternatively, by compelling some girls to stay until age 16, it may cause others that would have left at 16 to leave at 17 (e.g.,

to signal ability). We document that the 1972 ROSLA had no such effects and conclude that this particular concern is not a valid one. We discuss other possible violations of the exclusion restriction below.

3.5 Results

In this section we report our main results. We begin by reporting estimates of the impact of the ROSLA on educational attainment. We then report estimates of the impact of the ROSLA on teen fertility and cumulative fertility.

3.5.1 Education

The impacts of the 1972 ROSLA on educational attainment were previewed in Figure 3.1 above. In Table 3.2, we further assess these impacts by reporting a series of estimates based on data from the Labor Force Survey, 1975-2002. The sample underlying these estimates is restricted to women born in the UK and residing in England or Wales at the time of the interview. Each entry in the table displays the discontinuity estimate, γ_1 , from estimation of an equation similar to (1).

The estimate reported in the first column of panel A of Table 3.2 is obtained from a regression of the fraction of the cohort exiting school by age 15—that is, before age 16—on a ROSLA dummy and a linear cohort trend interacted with the ROSLA dummy. The estimation sample includes 48 months of birth cohort data on either side of the ROSLA threshold (i.e., 96 months in total). This estimate captures the discontinuity apparent in 3.1. The estimates in columns (2) and (3) are obtained from similar regressions in which the dependent variable is the fraction of women leaving school by ages 16 and 17 respectively. The column (4) estimate is obtained from a regression in which the dependent variable is the average age of school exit.

The estimates reported in panel B are obtained from models that include month of birth dummies. These are intended to capture seasonality in educational attainment. To allow

the seasonality to differ across the pre- and post-ROSLA periods, the panel C estimates are obtained from models that interact these month dummies with the ROSLA dummy. Since the data are collapsed to the cohort level, regressions are weighted by cohort size and robust standard errors are reported. These estimates are numerically identical to those that would be obtained from individual-level regressions in which standard errors are clustered at the cohort level.

The estimate in column (1) of panel A implies that the ROSLA decreased the population of women exiting school before age 16 by around 25 percentage points. This implies fully one quarter of the female population received an additional year of schooling between ages 15 and 16 as a result of the reform. In contrast, the estimates in columns (2) and (3) imply that the same fraction of women continued to exit school by age 16 after the reform. This suggests that there were no spillovers to education beyond the period of compulsory schooling. As noted above, spillovers would have occurred if girls for whom the reform was binding were somehow induced to stay beyond the new compulsory age or if other girls were induced to stay longer, perhaps to differentiate themselves from the new mass of students leaving at 16. The column (4) estimates, which measure average age of school exit, are less precise than those reported in the other columns, but yield a similar insight: on average, the ROSLA increased educational attainment among women by a quarter of a year. Comparisons of panels A, B, and C reveal that the estimates are not particularly sensitive to seasonality controls. Estimates based on different data windows also yield similar results (not reported).

Figure 3.4 provides an illustration of the robustness of these education effects. This figure corresponds closely to the column (4) estimates. The data points represent month of birth cohorts and a vertical line indicates the ROSLA threshold. The fitted lines represent education levels predicted using the trends estimated in the models without controls for month of birth (i.e., those used to generate the panel A estimates).

In the fertility analysis that follows, we interpret our estimates as the average effect

of a year of education at age 15 for women who would not have received this education in the absence of the ROSLA (i.e., a local average treatment effect (Imbens and Angrist 1994)). While effects for this subset of “compliers” may be different from the population-average effects, they identify effects for a group that is both policy-relevant (i.e., affected by compulsory schooling reforms) and large. For example, this fraction is twice as large as the fraction of girls affected by the Norwegian reform analyzed by Black et al. (2008), and much larger than the fraction of adolescents affected by various state-level compulsory schooling reforms in the US. In later sections, we use the value -0.248 (from column (1), Panel C of Table 3.2) as the denominator in our Wald IV estimates of the fertility impacts of schooling.

3.5.2 Teen Fertility

We turn next to ROSLA impacts on teen fertility. Figures 3.5 and 3.6 plot fertility rates at ages 16 and 17 among cohorts on either side of the reform. These fertility rates are calculated according to age-last-birthday. Thus the age 16 rate captures births in the exact age range [16,17), and the age 17 rate captures births in the exact age range [17,18). Again, each point in the figure represents the outcome for a particular month-of-birth cohort. A discontinuity associated with the ROSLA can be seen in both figures.

The magnitude and significance of these ROSLA effects is explored in Table 3.3. This presents estimates for fertility over all teen years, 15 through 19.¹⁴ The only significant effects are found for fertility at age 16 and 17. Since these effects are somewhat sensitive to specification, it helps to view these estimates in the context of Figures 3.5 and 3.6. Our preferred estimates are those in panel B, which are based on models that include month

¹⁴Our data do not allow us to examine fertility at ages below 15. The administrative birth records begin in 1970, which means that the first cohort for whom we could calculate an age 14 birth rate is January 1956. This leaves us with too small a pre-reform sample to do meaningful estimation. However, little is lost by this limitation. Fertility at age 14 is a rare event, even in our dataset of the universe of births. Thus any age 14 estimates would have been associated with large standard errors and driven by just a handful of births.

of birth dummies to remove seasonality. The panel B cohort trends are those plotted in Figures 3.5 and 3.6.¹⁵ Even the smaller estimates in panel B correspond to large fertility effects at ages 16 and 17. For example, the age 17 effect in panel B (-.00209) corresponds to a 5.8% decrease in fertility at age 17. After scaling by the first stage effect of ROSLA on schooling (a 25 percentage point increase in the fraction of women attending school for another year at age 15), the implied Wald IV estimate is a 23% reduction in age 17 fertility. The age 16 effect, which is significant in panel A but not in panel B (p-value of 0.14), is of a similar magnitude. The coefficient on age 16 fertility in panel B corresponds to a 5.5% reduced form effect and a 22% Wald IV effect.

The Table 3.3 teen fertility estimates are based on a narrower data window than the Table 3.2 education estimates. These different windows are those suggested by a cross-validation procedure. Intuitively, the fertility window is narrower because, around the ROSLA threshold, fertility trends are more rapidly changing than education trends. In suggesting a narrower window, the cross-validation procedure implicitly takes account of this feature of the data. Estimates based on different data windows and polynomial specifications yield similar results. See Appendix Table 3.6.

Since the ROSLA kept girls in school between ages 15 and 16, and since ROSLA fertility effects are concentrated between ages 16 and 17, it is important to pinpoint the exact timing of the ROSLA fertility effects and, ultimately, the exact timing of the conception effects that precede them. In a later section we address this question by estimating teen fertility effects at narrower age ranges - for example between ages 16 and 16 and one month. These analyses also help to determine whether there were statistically significant ROSLA effects on fertility prior to age 17. In relation to this question, the evidence presented in the first three columns of Table 3.3 is somewhat inconclusive, despite the visual evidence in Figure 3.5.

¹⁵These trends are estimated with month of birth dummies, but only the trends, not the dummies, are plotted.

Columns (4) and (5) of Table 3.3 indicate that there were no ROSLA impacts on fertility at ages 18 and 19. Figure 3.7 provides visual evidence. Consistent with the estimates in Table 3.3, neither of the panels in Figure 3.7 (for fertility at ages 18 and 19) suggests a discontinuity at the ROSLA threshold.

There does, however, appear to be a significant change in the slope of the age-19 fertility trend near this threshold. Upon closer inspection this represents a calendar time effect occurring around 1976-1977. This time effect, which occurs in the age 19 plot among cohorts born in late 1957, is also apparent in the age 18 plot among cohorts born in late 1958, in the age 17 plot among cohorts born in late 1959, and in the age 16 plot among cohorts born in late 1960. Since the graphs plotted in Figure 3.3 reveal that both France and Germany shared the same trend break, it seems safe to conclude that this has nothing to do with the ROSLA.

3.5.3 Cumulative Fertility

The finding of no ROSLA effects on fertility at ages 18 and 19 extends to ROSLA effects on fertility rates in one-year age intervals up to 45 (see the Appendix Figure 3.13 for results by single year of age). For a more succinct summary of these effects, and to see how they contribute to effects on completed family size, we now focus on cumulative fertility.

To begin, Figure 3.8 plots the RD figure for cumulative fertility from ages 16 to 45 (inclusive). Cohort-specific cumulative fertility is often referred to as the cohort total fertility rate (TFR).¹⁶ This cumulative rate is constructed as the sum of each cohort's age-specific rates. Because of the nature of our births data, we do not observe completed fertility at the person level. However, cohort TFR would exactly equal the average completed family size found in individual data provided that all women in the cohort survived through their childbearing years. Since the probability of survival from 16 to 45 among these cohorts was

¹⁶Cohort TFR, which is derived from age-specific rates for a fixed cohort, differs from period TFR, which is derived from age-specific rates for a fixed point in time. For example, Figure 3.2 plots period TFR, which sums age-specific rates of different cohorts in order to give a summary measure of fertility in each year.

above 98%, we view our cumulative fertility measure as a good proxy for average completed family size.

Figure 3.8 suggests that there was no ROSLA effect on completed fertility. The scale of the vertical axis measures average births per cohort member, and the apparent zero effect of the reform is contained within a very narrow band of births per woman. Effect sizes and standard errors are presented in Table 3.4, which follows the same format as earlier tables. The reduced form effects implied by the graph and shown in column (5) of Table 3.4 are on the order of +0.002 children. Even after this reduced form effect is scaled by the ROSLA effect on education, it is only +0.008 children. To put this effect size in context, note that for cohorts of women born near the reform threshold, an additional year of education at age 16 is associated with a decline in total childbearing of 0.248 children in the cross section (see Geruso 2010). Our IV estimate is not only an order of magnitude smaller, but also takes the wrong sign, although these wrong-signed estimates are indistinguishable from zero.

The finding of no completed fertility effect is robust to alternative regression specifications. Our main estimates, reported in Table 3.4, are based on models that use an interacted linear specification over a 19-month window, with and without dummies for month of birth. For completeness, we also include a cross-validation figure (Appendix Figure 3.14) that plots 12 different estimates based on alternative data windows, and in no case is the coefficient even marginally significant. In all cases, the standard errors are small. Appendix Table 3.7 lists results from using alternative polynomial controls for cohort trends. Here too, the results are fully robust to alternative specifications.

In even the simplest economic models of fertility, fertility and educational attainment are governed by the same joint decision process. These depend, among other things, on preferences, family background and peer influences, all forces that researchers cannot easily measure. These results suggest that the strong correlation found between education and completed family size reflects the impacts of these “third variables” on both education and

family size.

At first glance, it may seem that our findings of large negative teen fertility effects and small completed fertility effects imply positive “catch up” effects beyond the teenage years. Yet this is not necessarily true. Instead, because teen fertility is such a small component of completed fertility, the confidence intervals on our fertility effects include both this catch-up story and the polar opposite story under which the teen fertility effects lead to a permanent (albeit tiny) reduction in completed fertility.

These possibilities can be seen in Figure 3.9, which aggregates information from 30 distinct RD estimates. Each point represents an estimate of the cumulative fertility effect from 16 up to the indicated age. The left panel displays levels, the right panel logs. Since the first point is degenerate, representing the cumulative effect from 16 to 16, it corresponds to the ROSLA effect on the age 16 fertility rate (from column (2) and panel B of Table 3.3). The figure shows that the cumulative effects begin at age 16 and then die out, as the averted teen births become swamped by the higher number of births at older ages. This is consistent both with teen effects being counteracted by post-teen effects and with teen effects generating permanent reductions in completed fertility.

3.6 Discussion

In this section we discuss various explanations for our two main findings: sizable ROSLA effects on fertility at ages 16 and 17 and negligible ROSLA impacts on completed fertility. Note first that there is no contradiction between these results. Even if the teen fertility effects were permanent, they would have a tiny effect on completed fertility, one well within the already tight confidence intervals associated with our estimates. That is because teen fertility is such a small component of completed fertility. Put differently, we cannot hope to determine whether the teen fertility effects are permanent or whether positive ROSLA effects on post-teen fertility offset negative ROSLA effects on teen fertility.

3.6.1 Teen Fertility

At first glance, the teen fertility reductions appear to follow the compulsory schooling period with a lag of about one year, suggesting a connection between the fertility effects and the new period of schooling itself. Therefore, we now examine the timing of fertility effects relative to the timing of the new compulsory schooling period in more detail.

This conjecture that the reduction in pregnancies carried to term is influenced by current school attendance is supported by the analysis presented in Figure 3.10. Each panel plots ROSLA impacts on fertility rates defined over a succession of three-month intervals; the different panels correspond to different regression specifications. The leftmost point on each plot corresponds to ages 15-15.25, the second point to ages 15.25-15.5 and so on. The associated confidence intervals are represented by the dotted lines around each of these points. The vertical dashed lines depict the new compulsory schooling period shifted forwards by nine months to indicate where to expect a reduction in births if conceptions were reduced during the new compulsory enrollment. For most sample members, this additional compulsory schooling occurred between exact ages 15.5 and 16.5.¹⁷ Although the correspondence is not exact, the figure shows that roughly speaking, the ROSLA effects are largest 9-12 months after the new schooling period. All three specifications in the figure yield similar results with regard to the timing of effects.

Fertility reductions that occurred 9-12 months after the new compulsory schooling period must have resulted from some combination of reduced conception and increased termination during the compulsory schooling period. Increased termination is a plausible hypothesis: abortion was legalized by the time the ROSLA was implemented and bringing a pregnancy to term might have been less desirable for girls with several months of compulsory education left to complete.¹⁸

¹⁷Unlike in the US, students in the UK wishing to leave school at the earliest permissible date could not do so immediately upon their 16th birthday. In most cases, they had to complete a school term that ended some time after their 16th birthday.

¹⁸Cohorts at the threshold of the reform turned 16 in 1973, when abortion rates among teens 16 and 17 in England and Wales were 15 and 18 abortions per thousand women, respectively. By comparison, the 2010

In fact, the data suggest that there were no ROSLA effects on the rate of termination. This can be seen in Figure 3.11, which plots abortion rates at ages 16 through 19. Figure 3.11 reveals no discernible effect on abortion rates of the 1972 ROSLA. Regression results (not presented) confirm the visual evidence in the plots, indicating no significant discontinuity at ages 16 through 19 under a variety of specifications. This stands in contrast to the birth rate effects at 16 and 17.

To explain the teen fertility effects then, we must explain why girls were less likely to become pregnant during the new compulsory schooling period. An obvious explanation is that they had fewer opportunities to become pregnant: less contact and less time with older men that they might have met in other contexts, such as the workplace. Aggregate statistics from England and Wales show that most marriages among teen girls in this time period (including almost all teen shotgun weddings) are made with men two or more years older. To shed light on this “incarceration” story in which schooling kept girls separated from older potential partners, it would be useful to know more about the types of people that ROSLA-affected girls associated with and the types of activities that they engaged in. In future work, we plan to explore the NCDS data with a view to gathering this type of evidence. For example, the NCDS includes data on work and partnership histories, and this explanation about peer groups will be less plausible if we find that many ROSLA-affected girls nonetheless worked part-time and/or partnered with older men.

There are several other factors that could explain why girls were less likely to become pregnant during the new schooling period, although none are entirely convincing. First, the extra schooling may have weakened girls’ preferences for motherhood. It is not hard to imagine why education might have such an effect, but it is hard to explain why this type of effect would be temporary. That is, why would it not reduce fertility at later ages? Second, the extra schooling may have provided girls with better information about birth control.

the pooled abortion rate among 16-17 year-olds was 19 abortions per thousand women. Thus, women at the threshold of the reform experienced their teen years when utilization of abortions had already climbed to today’s levels.

Again though, one would expect girls to retain this informational advantage beyond the new compulsory schooling period, resulting in reduced fertility at later ages, including at 18 and 19.

3.6.2 Completed Fertility

Given the strong negative correlation between education and fertility often found in observational data, it is somewhat surprising that we find no evidence of an education effect on completed fertility. On face value, this also contradicts some theories of the education-fertility relationship, which predict that education reduces fertility via its impact on wage rates and hence on the opportunity cost of raising children.

It is easy to explain the difference between our results and least squares estimates of education's impact on total fertility that suggest a strong negative relationship. OLS estimates are unlikely to control for all of the factors that might generate a spurious correlation between education and fertility. This is particularly true since many of these are hard to measure, such as women's attitudes to their role in society.

In relation to theory, note that if children are a normal good, then other things equal, we would expect more-educated women to demand more of them. Willis (1973) acknowledges this, but argues that these positive effects are likely smaller than the negative effects operating via the opportunity cost mechanism. In the context of his model, this conclusion seems reasonable. This model does, however, impose some strong assumptions. Once these are relaxed, the positive income effects may be as strong as the negative opportunity cost (substitution) effects.

One important assumption is that women must choose between working and raising children. An alternative formulation might have women facing a labor market penalty for raising children, where this penalty varies across time, across countries and across occupations. If public policies and work practices have become more family-friendly over time, this penalty—and thus the strength of the opportunity cost mechanism—will have

weakened.

A second important assumption of classic fertility models is that education does not affect family formation: two-person households form exogenously and education levels influence decision-making within the household. In practice, it seems that women’s education might exert a strong influence on whether and to whom they get married. Becker (1981) has since formalized some of these ideas and there is now good evidence to suggest that more-educated women are more likely to marry more-educated and higher-earning men Lefgren and McIntyre (2006). From the perspective of subsequent fertility, this is likely to exert a second income effect, such that the combined income effects could either cancel out or dominate any opportunity cost effects.

To provide some UK evidence on this question, we commissioned a Census extract that contains details of the education levels of both partners in all marriages recorded in the 1991 Census. First, we estimated ROSLA effects on the probability of being married. Our estimates suggested that the ROSLA had no impact on the probability of being married. Next, we looked for ROSLA impacts on the partner’s education level. The results can be seen in Figure 3.12, which plots the highest educational qualifications attained by the male spouse. As in previous figures, it is the date of birth of each female cohort that defines the horizontal axis. The left panel of Figure 3.12 plots fraction of each cohort of women who married men with no qualifications. The right panel plots the fraction who married men with some O level qualifications, which would be based on national tests at the end of grade 10. The figure shows a jump across the reform. These effects are explored in more detail in Table 3.5. Column (1) shows that women affected by ROSLA were less likely to marry men with no qualifications. Columns (2) through (5) indicate that they instead married men with some O level qualifications.

These effects on spousal education levels are fairly large. For example, the effect on “no qualifications,” once scaled by the first stage, implies that an extra year of compulsory schooling reduces by five percentage points the probability of marrying a man with no

qualifications. Given conventional estimates of the rate of return to schooling, and assuming that the difference between no qualifications and some qualifications corresponds to a year of schooling, this spousal education channel could lead to important effects of women's education on household income.

Since the reform affected men and women, some of this spousal education effect will result from ROSLA-affected women being more likely to marry ROSLA-affected men, irrespective of their preferences for spousal education or opportunities to meet more-educated potential husbands. Policies that change women's education levels without changing those of men would likely have weaker effects. Nevertheless, theory and evidence suggests education effects on spousal education and it is interesting to see those effects in this context.

3.7 Conclusions

The problem of teen fertility is one that attracts a lot of attention. Discussions of how to reduce it have focused on sex education, the availability of contraception and the way the welfare system treats teenage mothers. In this paper we have shown that it might be reduced by policies that result in girls spending more time in education. It is not clear what drives this effect, although a leading explanation is that schooling has a type of incarceration effect similar to that documented by Jacob (2003) in relation to crime.

The decline in completed fertility is related problem that policy-makers have grappled with. Education does not feature in this policy discussion because, if anything, the conventional wisdom is that more-educated women have fewer children. Our analysis suggests that there is no causal relationship between education and completed fertility in the modern, developed country context we examine: Our estimates are statistically indistinguishable from zero and our confidence intervals restrict the possible effects to a small range.

Our findings are important because policy-makers in both the UK and US are considering measures that would result in girls spending more time in school. For example, the UK Coalition Government has reaffirmed the decision made by the previous Labour

Government to increase the compulsory schooling age to 17 by 2013 and 18 by 2015. Our results suggest that these measures could reduce teen fertility. Moreover, they suggest that these reductions could be achieved with little or no change in completed fertility. If the costs of teen fertility are as high as analysts claim, then these policies will generate important social benefits beyond those typically considered along dimensions such as crime and the labor market.

Data Source	Sample	Primary Variables	Notes
UK Labour Force Survey (LFS)	Sample of households in England and Wales 1973-2002	age left full-time education	individual-level
Live Birth Records Extract	100% sample of births in England and Wales 1970-2008	live births	counts by mother's and child's dob; not an individual-level panel
Abortions Records Extract	100% sample of teen abortions in England and Wales 1970-1979	abortions	counts by mother's dob and date of procedure
UK Census Extracts	100% sample 1991 and 2001	marital status, partner's age, partner's education	
National Child Development Study (1958 Cohort)	100% sample of children born one week in March 1978	activities in school, partnership histories, activities out of school	one of the first post-ROSLA Cohorts; forthcoming in future drafts

Table 3.1: Data sources used in the study.

Effects of ROSLA on Age of Exit from School

	left school by age 15 (1)	left school by age 16 (2)	left school by age 17 (3)	age left full-time ed (4)
Panel A: Linear cohort trend interacted with threshold				
discontinuity	-0.270*** (0.0191)	-0.0246* (0.00975)	-0.0105 (0.00677)	0.305*** (0.0623)
cohorts	96	96	96	96
Panel B: Month of birth indicators included				
discontinuity	-0.244*** (0.00982)	-0.0117 (0.00714)	-0.00333 (0.00476)	0.235*** (0.0357)
cohorts	96	96	96	96
Panel C: Month of birth indicators interacted with threshold included				
discontinuity	-0.248*** (0.0120)	-0.0129 (0.0103)	-0.00198 (0.0104)	0.255* (0.107)
cohorts	96	96	96	96

Table 3.2: Effects on the age of school-exit of the 1972 ROSLA, which raised the compulsory schooling age to 16. Each column and row is the discontinuity estimate from a separate regression. Column headers list the dependent variable. Panel A includes linear controls for cohort trends, interacted with the threshold to allow different pre- and post-reform trends. Panel B adds dummies for month of birth to capture seasonality. Panel C interacts the month dummies with the threshold to allow this seasonality to differ pre- and post-reform. Column 1 shows that ROSLA decreased the population of girls exiting school before age 16 by around 25 percentage points. There was little spillover to educational attainment beyond age 16, as columns 2 and 3 indicate no significant fraction of girls were induced to stay on beyond age 16. Data come from UK Labour Force Surveys, 1975-2002. The sample is restricted to women born in the UK and resident in England or Wales at the time of the interview. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effects of ROSLA on Teen Fertility

	age 15 fertility rate (1)	age 16 fertility rate (2)	age 17 fertility rate (3)	age 18 fertility rate (4)	age 19 fertility rate (5)
Panel A: Linear cohort trend interacted with threshold					
discontinuity	-0.000227 (0.000186)	-0.00182* (0.000709)	-0.00313*** (0.000766)	-0.00133 (0.00122)	-0.000168 (0.00145)
cohorts	38	38	38	38	38
Panel B: Month of birth indicators included					
discontinuity	-0.000160 (0.000252)	-0.000867 (0.000568)	-0.00209** (0.000718)	-0.000753 (0.00106)	0.000412 (0.00129)
cohorts	38	38	38	38	38

Table 3.3: Fertility effects of the 1972 ROSLA, showing a significant impact on fertility outcomes at 16 and 17. Each column and row is the discontinuity estimate from a separate regression. Column headers list the dependent variable. Panel A includes linear controls for cohort trends, interacted with the threshold to allow different pre- and post-reform trends. Panel B adds dummies for month of birth to capture seasonality. Data come from live birth records supplied by the UK Office of National Statistics, containing all births in England and Wales to mothers born in England, Wales, and Scotland, 1970-2008. Each observation is a cohort-cell mean. Regressions are weighted by cohort sizes. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effects of ROSLA on Cumulative Fertility

	Cumul. fertility 16-25 (1)	Cumul. fertility 16-30 (2)	Cumul. fertility 16-35 (3)	Cumul. fertility 16-40 (4)	Cumul. fertility 16-45 (5)
Panel A: Linear cohort trend interacted with threshold					
discontinuity	-0.00631 (0.00733)	-0.000526 (0.00725)	0.00220 (0.00595)	0.00367 (0.00584)	0.00358 (0.00578)
cohorts	38	38	38	38	38
Panel B: Month of birth indicators included					
discontinuity	-0.000387 (0.00552)	0.00311 (0.00588)	0.00315 (0.00562)	0.00411 (0.00579)	0.00370 (0.00585)
cohorts	38	38	38	38	38

Table 3.4: Effects of the 1972 ROSLA on cumulative fertility at various ages. Each column and row is the discontinuity estimate from a separate regression. There are no significant effects for any of these cumulative measures, indicating the ROSLA had no impact on family size from age 25 to 45. Column headers list the dependent variable. Panel A includes linear controls for cohort trends, interacted with the threshold to allow different pre- and post-reform trends. Panel B adds dummies for month of birth to capture seasonality. Cumulative fertility at each age x is the sum of age-specific fertility rates from 16 to x , inclusive. Data come from live birth records supplied by the UK Office of National Statistics, containing all births in England and Wales to mothers born in England, Wales, and Scotland, 1970-2008. Each observation is a cohort-cell mean. Regressions are weighted by cohort sizes. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Husband's Highest Qualifications

	No Quals (1)	O-Levels 1 to 4 (2)	O-Levels 5 or more (3)	O-Levels any number (4)	A-Levels or higher (5)
Panel A: Linear cohort trend interacted with threshold					
discontinuity	-0.0122*** (0.00223)	0.00588*** (0.00158)	0.00638*** (0.00145)	0.0123*** (0.00159)	0.00360 (0.00250)
cohorts	72	72	72	72	72
Panel B: Month of birth indicators included					
discontinuity	-0.0123*** (0.00191)	0.00578** (0.00184)	0.00665*** (0.00166)	0.0124*** (0.00182)	0.00278 (0.00208)
cohorts	72	72	72	72	72

Table 3.5: Effects on husband's qualifications. The window of estimation is 72 months–36 on either side—and was chosen via a cross-validation function. Panel A includes linear controls for cohort trends, interacted with the threshold to allow different pre- and post-reform trends. Panel B adds dummies for month of birth to capture seasonality. Data come from the 2001 Census. Each observation is a cohort-cell mean. Regressions are weighted by cohort sizes. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Age 17 Fertility: Robustness to Alternative Specifications

	bandwidth on either side of threshold			
	17	19	21	23
1st order	-0.00197 0.00086	-0.00200 0.00079	-0.00231 0.00077	-0.00239 0.00076
1st order interacted	-0.00205 0.00080	-0.00209 0.00072	-0.00240 0.00070	-0.00251 0.00064
2nd order	-0.00196 0.00076	-0.00201 0.00069	-0.00233 0.00069	-0.00243 0.00062
2nd order interacted	-0.00138 0.00167	-0.00174 0.00132	-0.00117 0.00111	-0.00133 0.00098
3rd order	-0.00151 0.00142	-0.00186 0.00108	-0.00139 0.00091	-0.00163 0.00082
3rd order interacted	-0.00163 0.00204	-0.00144 0.00188	-0.00161 0.00181	-0.00138 0.00157
optimal poly order (AIC)	2nd	2nd	3rd	3rd

Table 3.6: Each column and row is the discontinuity estimate from a separate regression with the indicated specification. The dependent variable is the age 17 fertility rate. Month of birth indicators are included in all regressions to control for seasonality. Column headers list the number of cohorts included in the regression on each side of the reform. Row headers list the polynomial order. Interacted specifications allow the polynomial coefficients to differ across the threshold. The optimal polynomial order is determined by Akaike's information criterion. Regressions are weighted by cohort sizes. Robust standard errors in parentheses.

Total Fertility: Robustness to Alternative Specifications				
	bandwidth on either side of threshold			
	17	19	21	23
1st order	0.00562 (0.00590)	0.00315 (0.00601)	0.00117 (0.00611)	-0.00169 (0.00633)
1st order interacted	0.00618 (0.00544)	0.00370 (0.00585)	0.00168 (0.00584)	-0.00140 (0.00629)
2nd order	0.00558 (0.00540)	0.00323 (0.00582)	0.00130 (0.00582)	-0.00162 (0.00632)
2nd order interacted	-0.01142 (0.01154)	0.00218 (0.01001)	0.00923 (0.00917)	0.01498 (0.00837)
3rd order	-0.00736 (0.00856)	0.00353 (0.00810)	0.00803 (0.00775)	0.01086 (0.00746)
3rd order interacted	-0.00300 (0.01553)	-0.00915 (0.01589)	-0.01656 (0.01677)	-0.01451 (0.01473)
optimal poly order (AIC)	3rd	1st-interact	1st-interact	3rd-interact

Table 3.7: Each column and row is the discontinuity estimate from a separate regression with the indicated specification. The dependent variable is total fertility from age 16 to 45. Month of birth indicators are included in all regressions to control for seasonality. Column headers list the number of cohorts included in the regression on each side of the reform. Row headers list the polynomial order. Interacted specifications allow the polynomial coefficients to differ across the threshold. The optimal polynomial order is determined by Akaike's information criterion. Regressions are weighted by cohort sizes. Robust standard errors in parentheses.

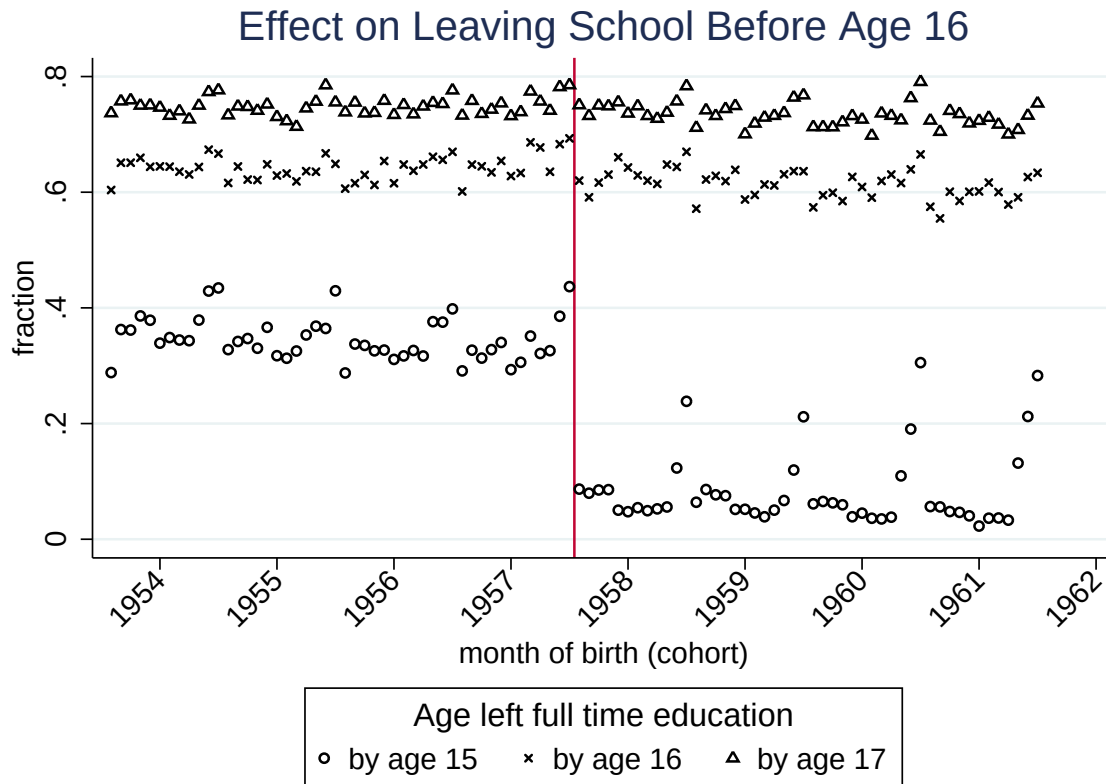


Figure 3.1: The impact of the 1972 ROSLA on educational attainment. Each point represents a cohort of women sharing the same month of birth. The vertical line indicates the ROSLA threshold, with the first post-reform cohort born September 1957. The three trends in the figure trace the fraction of women who left school by age 15, by age 16, and by age 17. Age 15 corresponds to grade 9 and age 16 corresponds to grade 10. The discontinuity at the threshold implies that around 25% of the female population received an additional year of schooling between ages 15 and 16 as a result of the reform. Data come from the UK Labor Force Survey, pooled 1975-2002 and restricted to women who were born in the UK and resident in England or Wales at the time of the interview. Respondents were asked the age at which they first left continuous, full-time education.

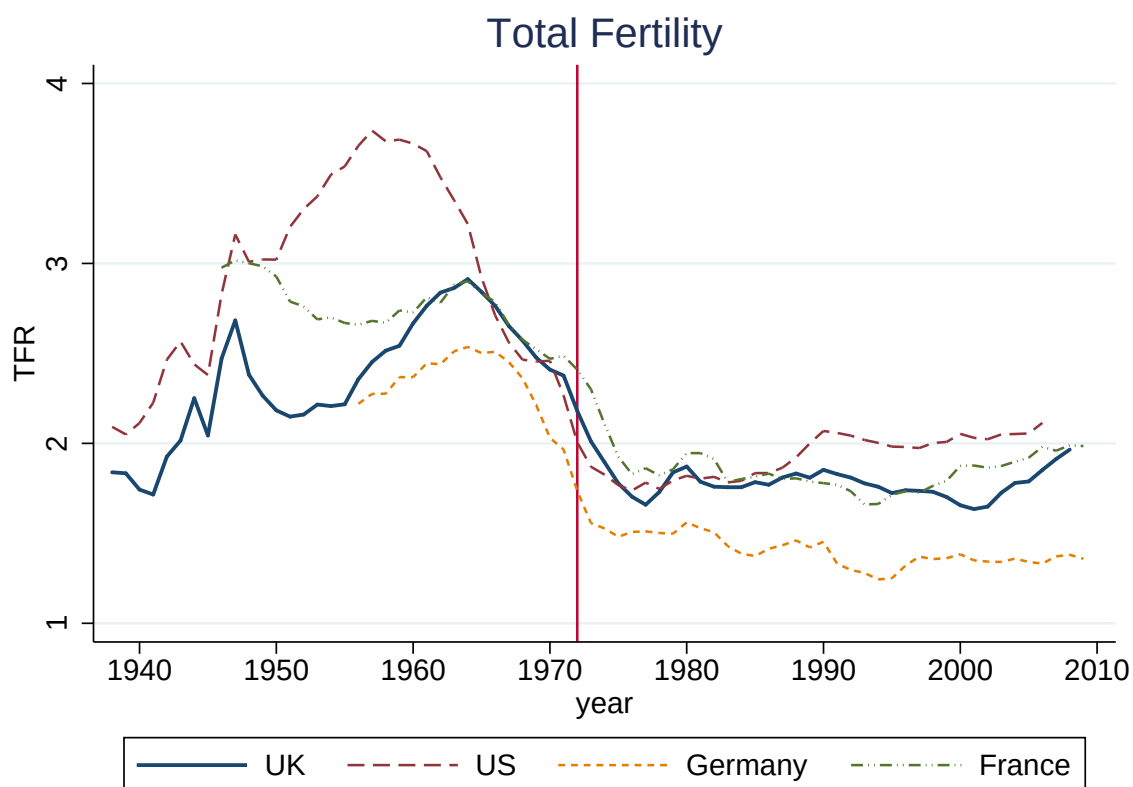


Figure 3.2: Total fertility rates in the UK, US, France, and Germany. The vertical line indicates the date of the 1972 ROSLA. The raising of the school leaving age in 1972 came at the tail end of a demographic transition among industrialized nations. Women just affected by the reform experienced their childbearing years (roughly 1972 through 2002) in a period of low, stable fertility. UK data come from the UK Office of National Statistics. US, French, and German data come from the Human Fertility Database.

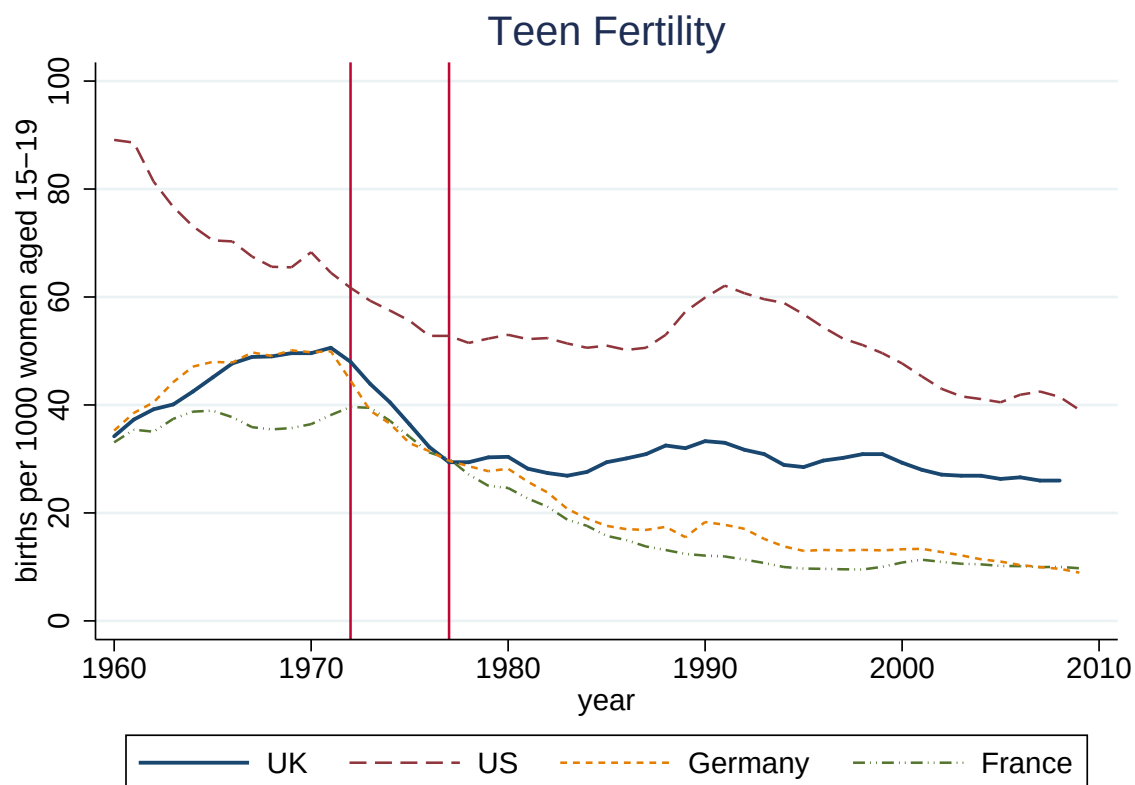


Figure 3.3: Teen fertility rates in the UK, US, France, and Germany. The vertical axis counts births per 1000 women aged 15 to 19. Comparisons of trends in the UK, US, France and Germany indicates that the steep downward trend in teen fertility through the mid-1970s was not a phenomenon limited to the UK, which experienced an educational reform in 1972. The two vertical lines delimit the years during which women born on the cusp of the educational reform were aged 15 to 19 (1972 to 1977). UK data come from the UK Office of National Statistics. US data come from the US National Center for Health Statistics. French and German data come from the Human Fertility Database.

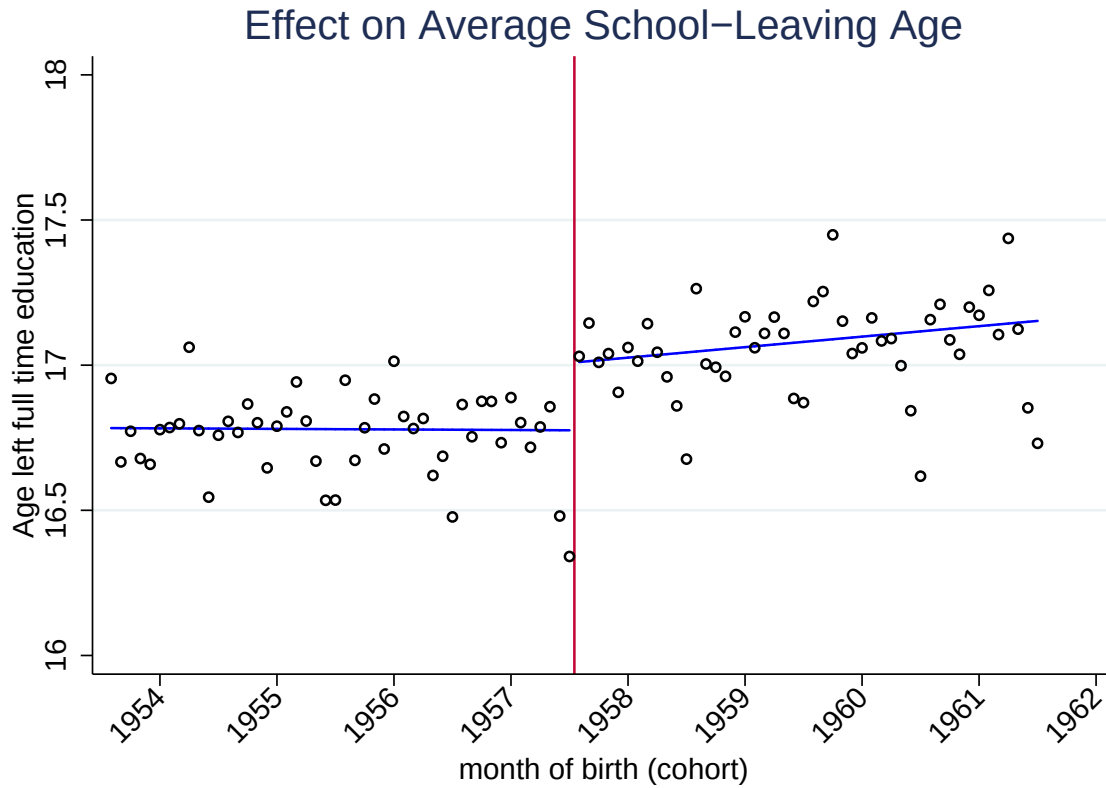


Figure 3.4: The impact of the 1972 ROSLA on average age of school exit. The discontinuity at the threshold implies that average age of school exit increased by around a quarter year as a result of the reform. Each point represents a cohort of women sharing the same month of birth. The vertical line indicates the ROSLA threshold, with the first post-reform cohort born September 1957. The fitted solid lines control for cohort trends, and are estimated over two 48-month windows on each side of the reform. Data come from the UK Labor Force Survey, pooled 1975-2002 and restricted to women who were born in the UK and resident in England or Wales at the time of the interview. Respondents were asked the age at which they first left continuous, full-time education.

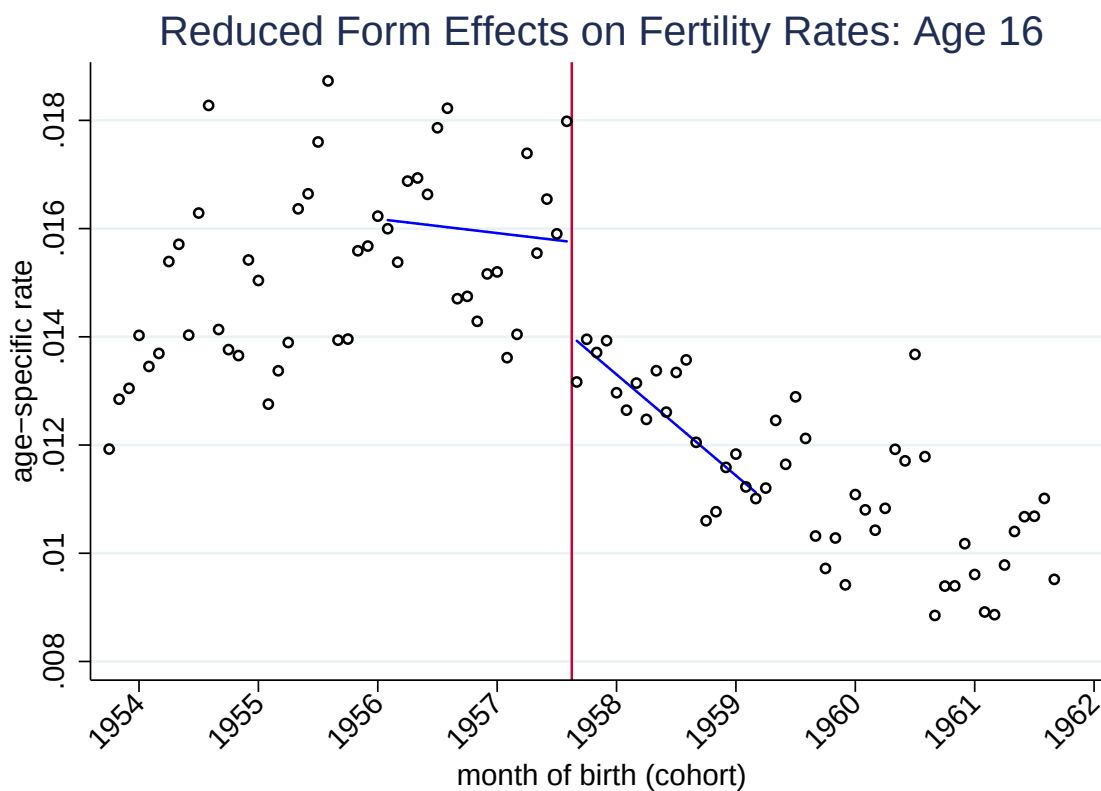


Figure 3.5: The reduced form impact of the 1972 ROSLA on fertility at age 16. The discontinuity at the threshold corresponds to roughly a 6% reduction in fertility. Once this reduced form effect is scaled by the first stage, the resulting Wald IV estimate indicates that among women kept in school an additional year by the reform, age 16 fertility was reduced by more than 20%. Each point represents a cohort of women sharing the same month of birth. The fitted solid lines control for cohort trends, and are estimated separately in a 19-month window on each side of the reform. The fertility rate at age x is defined as the number of children born to women aged x , divided by the number of women in the cohort alive at age x .

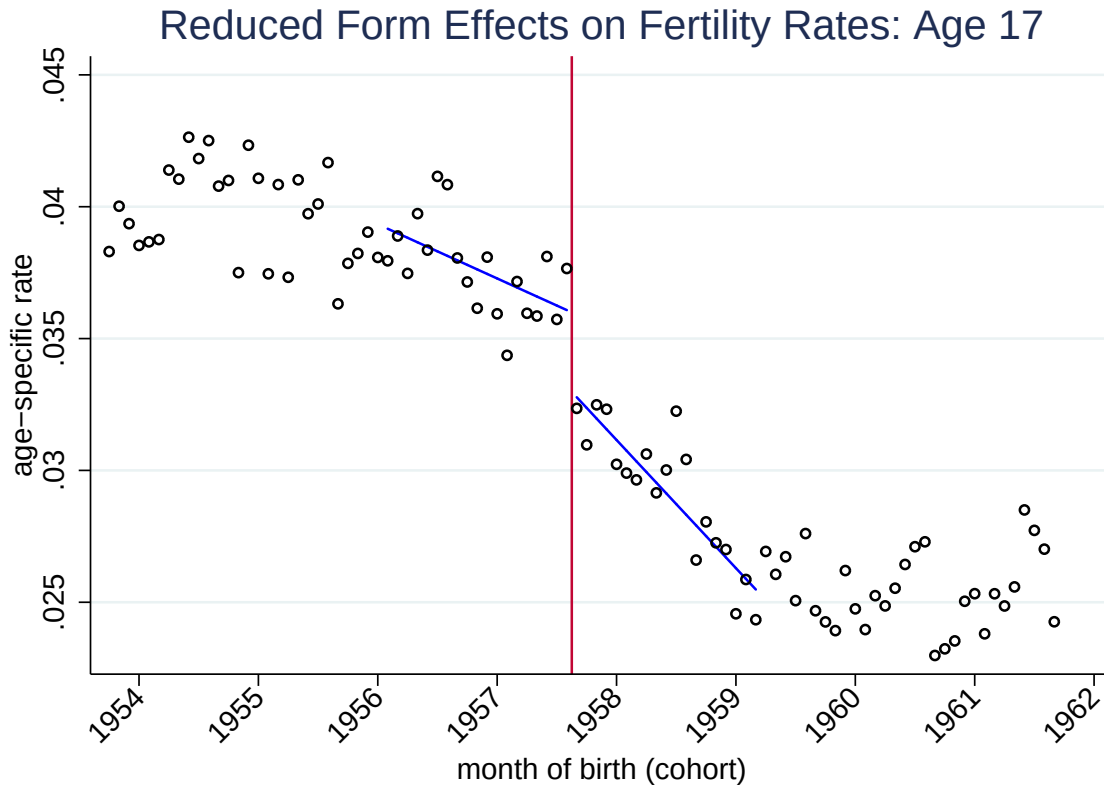


Figure 3.6: The reduced form impact of the 1972 ROSLA on fertility at age 17. The discontinuity at the threshold corresponds to roughly a 6% reduction in fertility. Once this reduced form effect is scaled by the first stage, the resulting Wald IV estimate indicates that among women kept in school an additional year by the reform, age 17 fertility was reduced by more than 20%. Each point represents a cohort of women sharing the same month of birth. The fitted solid lines control for cohort trends, and are estimated separately in a 19-month window on each side of the reform. The fertility rate at age x is defined as the number of children born to women aged x , divided by the number of women in the cohort alive at age x .

Reduced Form Effects on Fertility Rates: 18 and 19

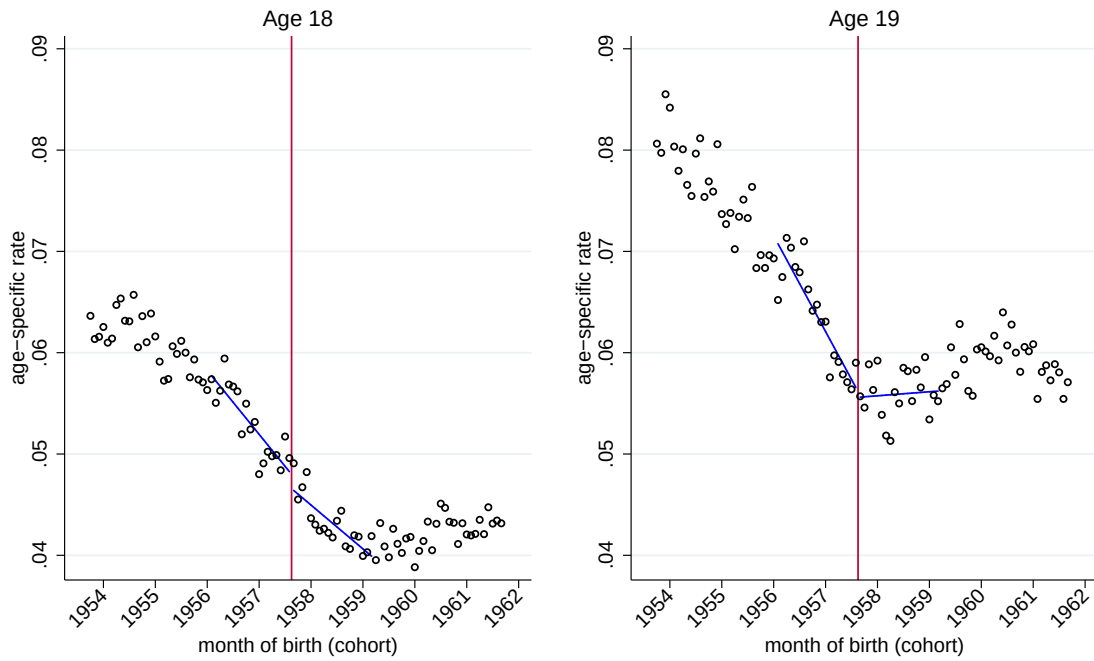


Figure 3.7: The reduced form impact of the 1972 ROSLA on fertility at ages 18 and 19. The lack of discontinuity at the thresholds in both panels suggests the 1972 ROSLA had no effects on fertility at ages 18 and 19. Each point represents a cohort of women sharing the same month of birth. The fitted solid lines control for cohort trends, and are estimated separately in a 19-month window on each side of the reform. The fertility rate at age x is defined as the number of children born to women whose age-last-birthday was x , divided by the number of women in the cohort alive at age x .

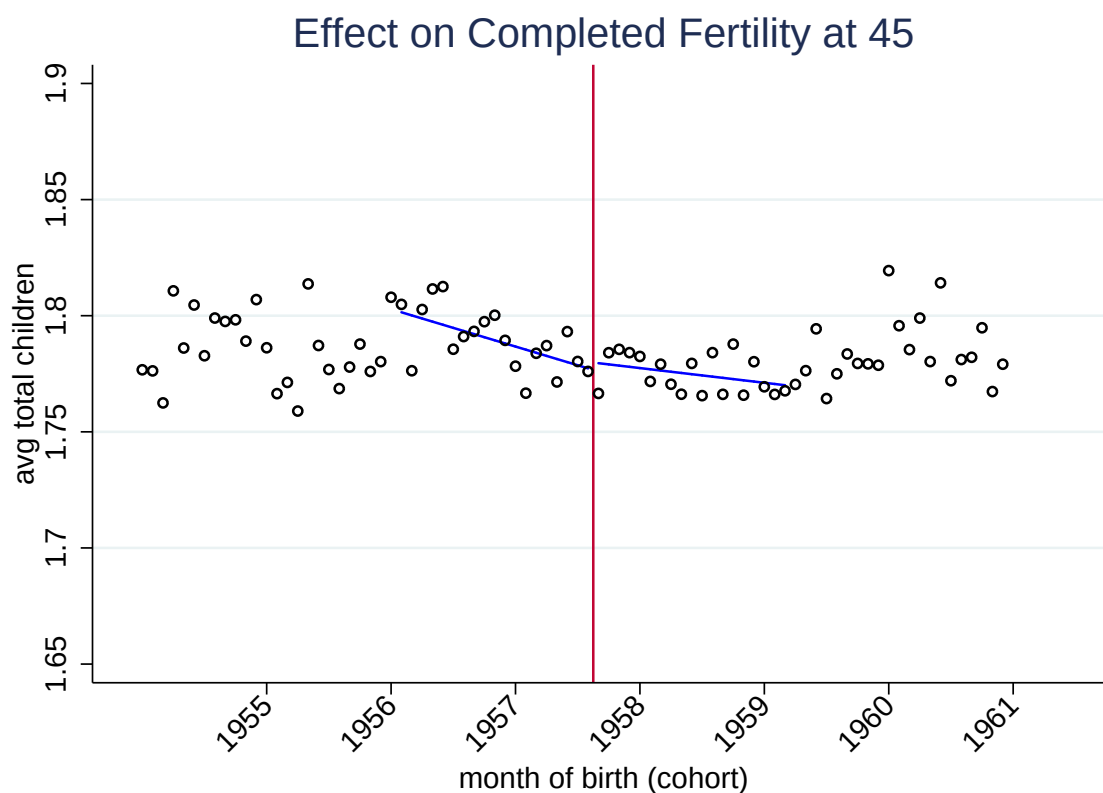


Figure 3.8: The reduced form impact of the 1972 ROSLA on cumulative fertility to age 45. The implied discontinuity and its associated standard error rule out even small effects of the reform on completed fertility. Each point represents a cohort of women sharing the same month of birth. The fitted solid lines control for cohort trends, and are estimated separately in a 19-month window on each side of the reform. The completed fertility rate at age 45 is defined as the sum of age specific rates from 16 to 45.

Evolution of Cumulative Fertility Effects

Estimation specification: linear trends interacted with threshold
19 month window either side, mob dummies included

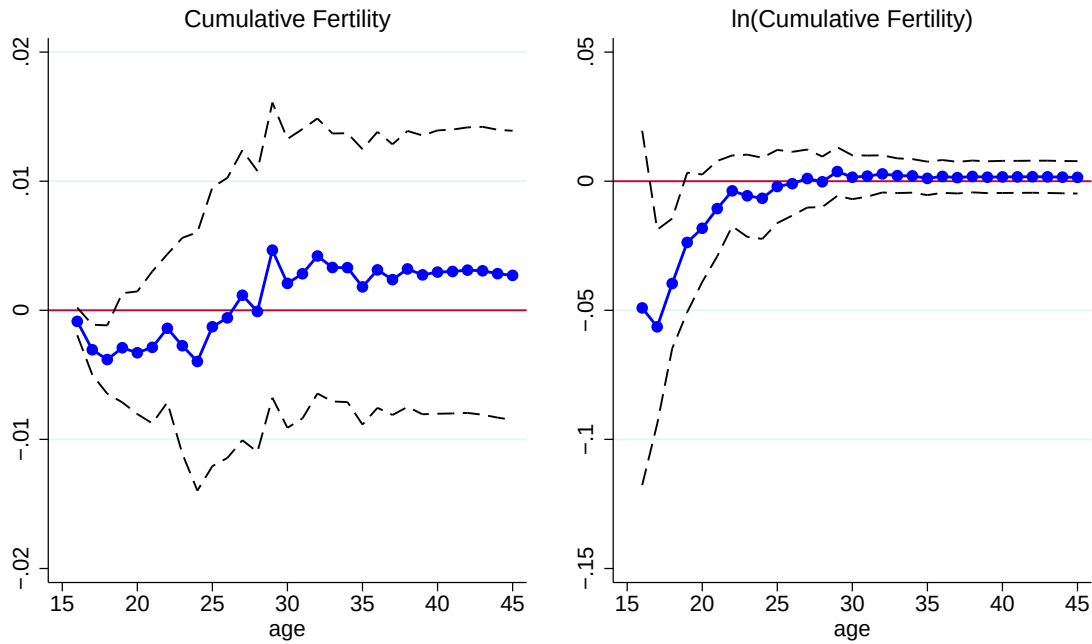


Figure 3.9: Effects of the 1972 ROSLA on cumulative fertility. Each point in the plots represents a distinct RD estimate of the reduced form effect of ROSLA on cumulative fertility from 16 up to and including the indicated age. Pointwise 95% confidence intervals are plotted as dotted lines. The left panel plots level effects, which are measured in counts of children per woman. The right panel plots effects on log of cumulative fertility. The left panel shows that the confidence interval generally rules out effects larger than +.015 births per woman and smaller than -.010 birth per woman, and is significant only at ages 17 and 18. The right panel, which scales effects as percentage changes, shows that large percentage effects occur in the teen years and then die out. Large teen effects are consistent with very small effects on cumulative fertility at age 45 because the levels of the teen effects are small compared to completed fertility. Cumulative fertility at each age x is the sum of age-specific fertility rates from 16 to x , inclusive. Since the first point in each panel is degenerate, representing the cumulative effect from 16 to 16, it corresponds exactly with the discontinuity estimate of the reform on the age 16 fertility rate.

Effects over teen years

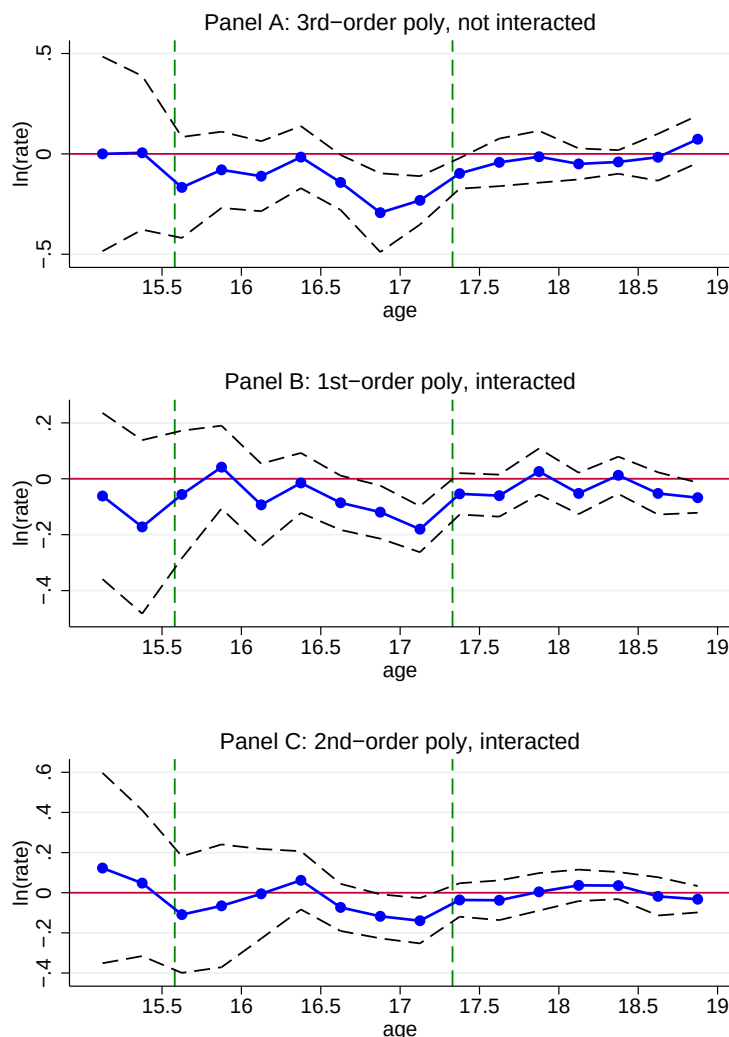


Figure 3.10: Detailed timing of birth effects. Fertility rates are calculated over quarter-years of age. Each point represents a distinct RD estimate of the effect of ROSLA on fertility in a quarter-year age interval. The first point represents the RD estimate of the reform on births between ages 15.00 and 15.25, the second point represents the RD estimate of the reform on births between ages 15.25 and 15.50, and so on. The vertical dotted lines denote the time during which the effected cohorts were newly in school, shifted right by 9 months. Pointwise 95% confidence intervals are plotted as dotted lines. The three panels show results under alternative specifications of the trend control. The window of estimation of the trend control was determined using a cross-validation procedure described in the text.

Abortion Rates

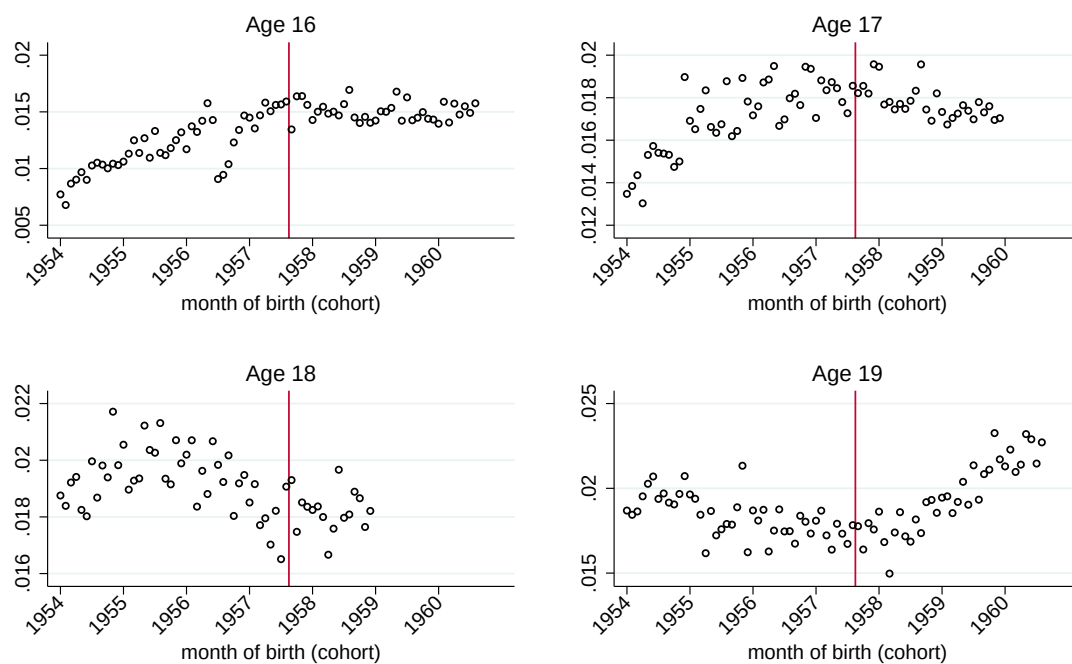


Figure 3.11: The reduced form impact of the 1972 ROSLA on abortions. The lack of discontinuity at the thresholds across all panels indicates that the 1972 ROSLA had no effects on abortion rates at ages 16 through 19. The abortion rate at age x is defined as the number of abortions performed on women whose age-last-birthday was x , divided by the number of women in the cohort alive at age x .

Husband's Highest Qualifications

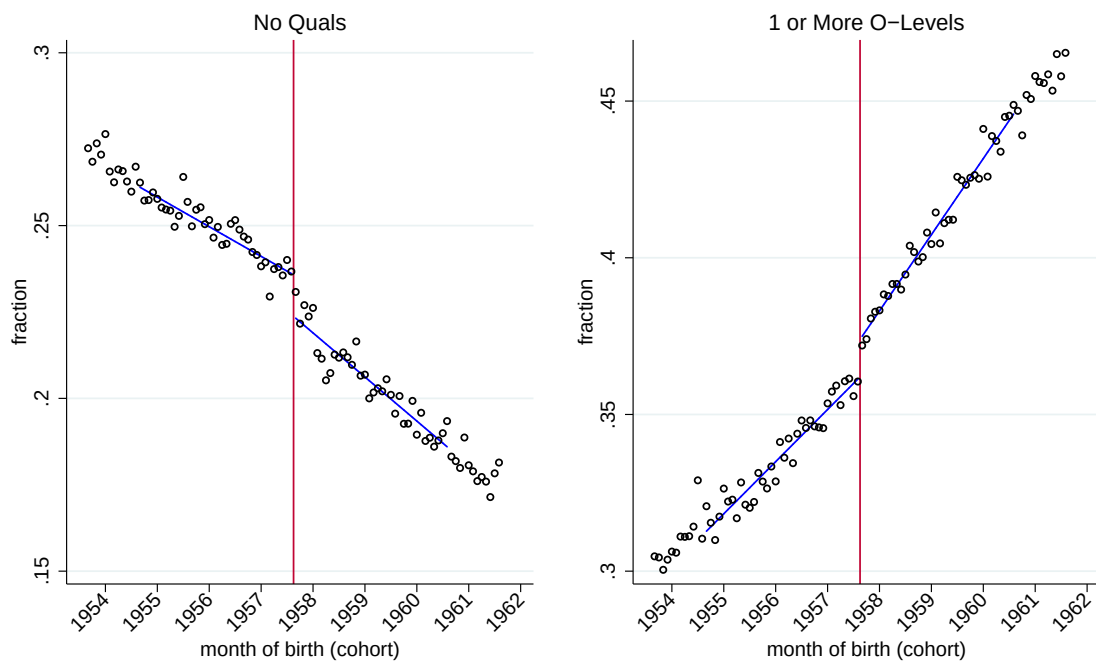


Figure 3.12: Effects on husband's educational qualifications. The horizontal axis is the wife's date of birth. Data come from the 2001 Census. Each observation is a cohort-cell mean.

Evolution Age-Specific Fertility Effects

Estimation specification: linear trends interacted with threshold
19 month window either side, mob dummies included

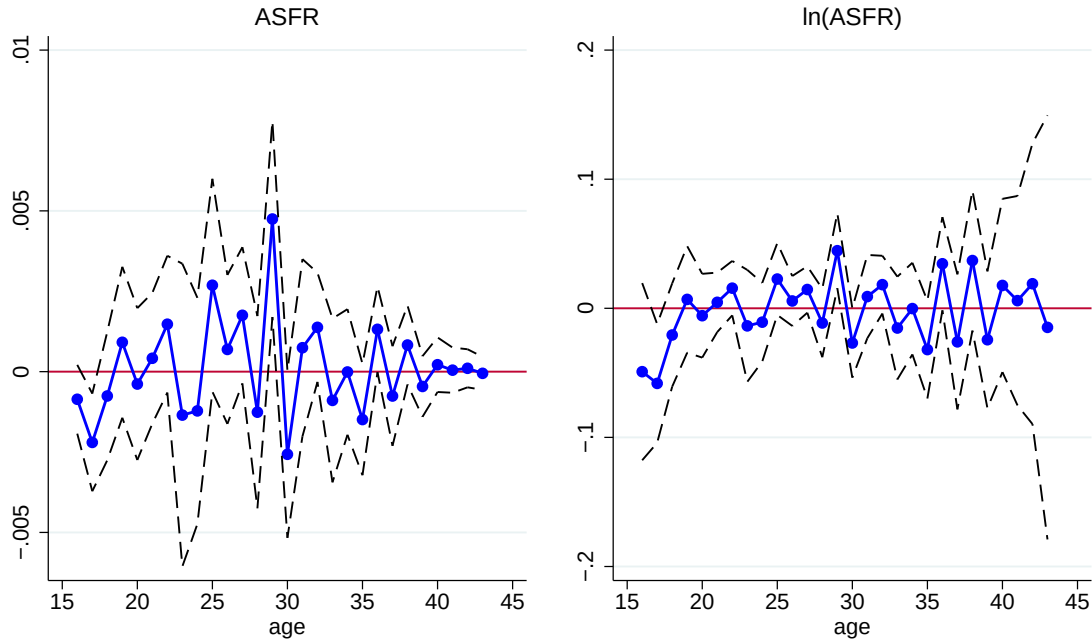


Figure 3.13: Effects of the 1972 ROSLA on age specific fertility at each age. Each point in the plots represents a distinct RD estimate of the reduced form effect of ROSLA on an age-specific fertility rate. Pointwise 95% confidence intervals are plotted as dotted lines. The left panel plots level effects on rates. The right panel plots effects on log of of these rates.

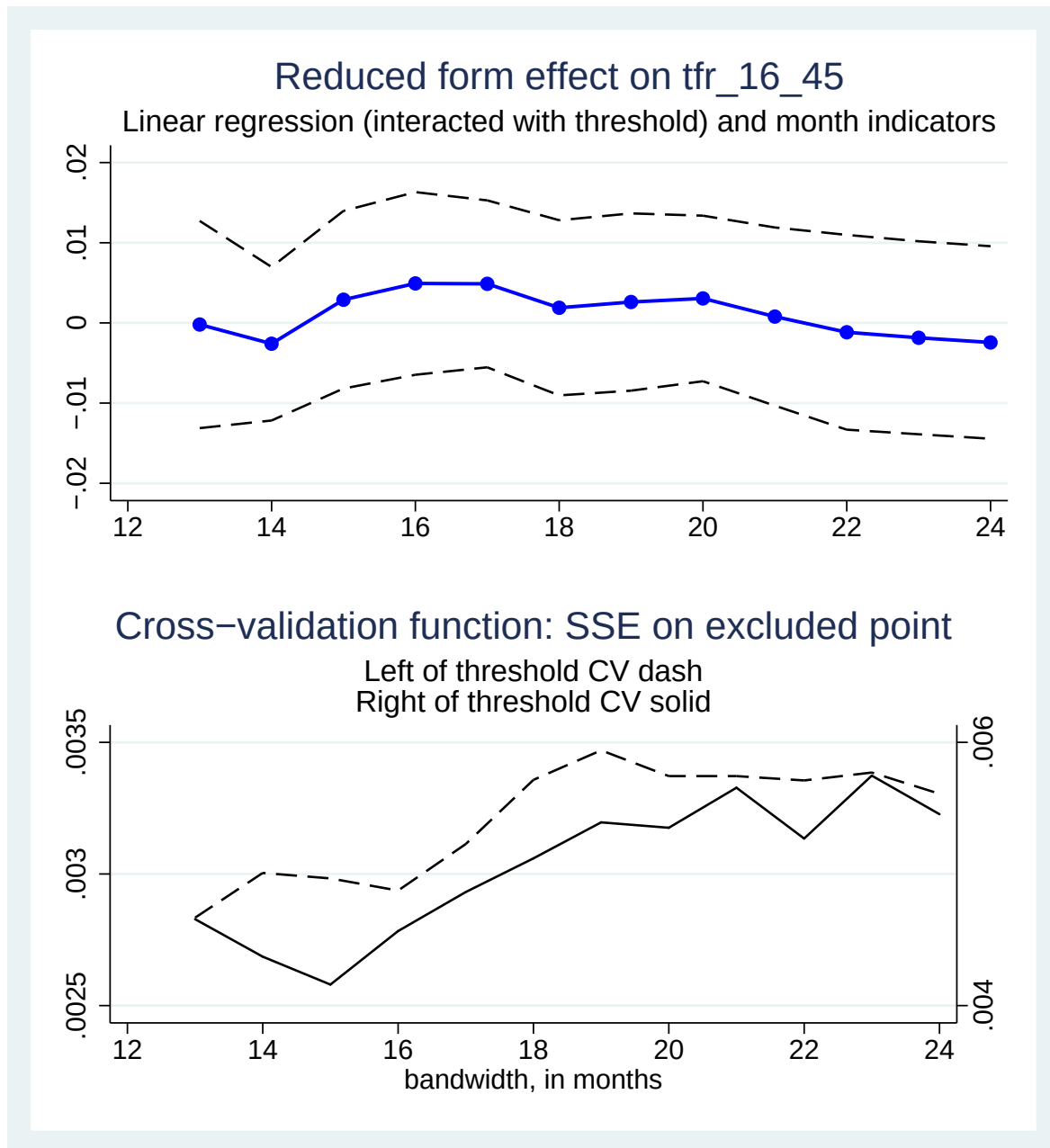


Figure 3.14: ROSLA's effect on TFR under different specifications of the cohort trends. The top panel plots RD estimates for the effect of ROSLA on cumulative fertility at age 45 for a range of windows used to estimated the cohort trend. In all cases, the cohort trend is linear, estimated separately on either side of the reform, and includes month of birth dummies to capture seasonality. The bottom panel plots the sum of squared residuals from a cross-validation function calculated for the same set of trend specifications. The cross-validation procedure is described in detail in Lee and Lemieux (2010). Lower values of the function indicate a better fit of the trend.

Bibliography

- Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):pp. 488–500.
- Angrist, J. D. and Krueger, A. B. (1992). Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery. Working paper no. 4067, National Bureau of Economic Research.
- Arias, E., Rostron, B. L., and Tejada-Vera, B. (2010). United States Life Tables, 2005. *National Vital Statistics Reports*, 58(10).
- Arriaga, E. E. (1984). Measuring and Explaining the Change in Life Expectancies. *Demography*, 21(1):83–96.
- Arriaga, E. E. (1989). Changing Trends in Mortality Decline During the Last Decades. In Ruzicka, L., Wunsch, G., and Kane, P., editors, *Differential Mortality: Methodological Issues and Biosocial Factors*, pages 105–129. Clarendon Press.
- Ashcraft, A. and Lang, K. (2006). The Consequences of Teenage Childbearing. Working paper no. 12485, National Bureau of Economic Research.
- Becker, G. S. (1960). An Economic Analysis of Fertility. In *Demographic and Economic Change in Developed Countries*. National Bureau of Economic Research.
- Becker, G. S. (1981). Altruism in the Family and Selfishness in the Market Place. *Economica*, 48(189):pp. 1–15.

- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2008). Staying in the Classroom and out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births. *The Economic Journal*, 118(530):1025–1054.
- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4):436–455.
- Breierova, L. and Duflo, E. (2004). The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers? Working Paper 10513, National Bureau of Economic Research.
- Buchmueller, T. and DiNardo, J. (2002). Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut. *The American Economic Review*, 92(1):pp. 280–294.
- Bundorf, K. M., Levin, J., and Mahoney, N. (2008). Pricing, Matching and Efficiency in Health Plan Choice. *mimeo, Stanford University*.
- Carlin, C. and Town, R. (2008). Adverse Selection, Welfare and Optimal Pricing of Employer-Sponsored Health Plans. *mimeo, University of Minnesota*.
- Case, A., Lubotsky, D., and Paxson, C. (2002). Economic Status and Health in Childhood: The Origins of the Gradient. *The American Economic Review*, 92(5):1308–1334.
- Case, A. and Paxson, C. (2011). The Long Reach of Childhood Health and Circumstance: Evidence from the Whitehall II Study. *The Economic Journal*, 121(554):F183–F204.
- Chevalier, A., Viitanen, T. K., and Viitanen, T. K. (2003). The Long-run Labour Market Consequences of Teenage Motherhood in Britain. *Journal of Population Economics*, 16:323–343.
- Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. John Wiley & Sons, Inc.

- Clark, D. and Royer, H. (2010). The Effect of Education on Adult Health and Mortality: Evidence from Britain. Working Paper 16013, National Bureau of Economic Research.
- Cohen, A. and Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *The American Economic Review*, 97(3):pp. 745–788.
- Currie, J. and Moretti, E. (2003). Mother’s Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. *The Quarterly Journal of Economics*, 118(4):1495–1532.
- Cutler, D. M., Finkelstein, A., and McGarry, K. (2008). Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance. *The American Economic Review*, 98(2):pp. 157–162.
- Cutler, D. M. and Reber, S. J. (1998). Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection. *The Quarterly Journal of Economics*, 113(2):433–466.
- Das Gupta, P. (1993). *Standardization and Decomposition of Rates: A User’s Manual*. Washington, DC: U.S. Government Printing Office.
- Deaton, A. (2002). Policy Implications of the Gradient of Health and Wealth. *Health Affairs*, 21(2):13.
- Devereux, P. J. and Hart, R. A. (2010). Forced to be Rich? Returns to Compulsory Schooling in Britain. *The Economic Journal*, 120(549):1345–1364.
- DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5):1001–1044.
- Duflo, E., Dupas, P., and Kremer, M. (2011). Education, HIV and Early Fertility: Experimental Evidence from Kenya. Working paper, MIT.

- Ecob, R. and Smith, G. D. (1999). Income and Health: What Is the Nature of the Relationship? *Social Science & Medicine*, 48(5):693–705.
- Einav, L. and Finkelstein, A. (2011). Selection in Insurance Markets: Theory and Empirics in Pictures. *Journal of Economic Perspectives*, 25(1):115–38.
- Einav, L., Finkelstein, A., and Cullen, M. R. (2010a). Estimating Welfare in Insurance Markets Using Variation in Prices. *The Quarterly Journal of Economics*, 125(3):877–921.
- Einav, L., Finkelstein, A., and Levin, J. (2010b). Beyond Testing: Empirical Models of Insurance Markets. *Annual Review of Economics*, 2(1):311–336.
- Elo, I. and Preston, S. (1996). Educational Differentials in Mortality: United States, 1979–1985. *Social Science & Medicine*, 42(1):47–57.
- Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. *Journal of Political Economy*, 116(2):pp. 303–350.
- Farrell, P. and Fuchs, V. (1982). Schooling and Health: The Cigarette Connection. *Journal of Health Economics*, 1(3):217–230.
- Feldman, R. and Dowd, B. (2000). Risk Segmentation: Goal or Problem? *Journal of Health Economics*, 19(4):499 – 512.
- Finch, B. (2003). Early Origins of the Gradient: The Relationship Between Socioeconomic Status and Infant Mortality in the United States. *Demography*, 40(4):675–699.
- Finkelstein, A. and McGarry, K. (2006). Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market. *The American Economic Review*, 96(4):pp. 938–958.

- Geronimus, A. T. and Korenman, S. (1992). The Socioeconomic Consequences of Teen Childbearing Reconsidered. *The Quarterly Journal of Economics*, 107(4):1187–1214.
- Geruso, M. (2010). Education and Childbearing: What are the Links? CMPO Bulletin, University of Bristol.
- Goldin, C. and Katz, L. F. (2008). Mass Secondary Schooling and the State: The Role of State Compulsion in the High School Movement. In *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, pages 275–310. University of Chicago Press.
- Goldman, N. (1993). Marriage Selection and Mortality Patterns: Inferences and Fallacies. *Demography*, 30(2):189–208.
- Goodman, A., Kaplan, G., and Walker, I. (2004). Understanding the Effects of Early Motherhood in Britain: The Effects on Mothers. Discussion paper no. 1131, Institute for the Study of Labor (IZA) Research Paper Series.
- Handel, B. (2010). Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts. *mimeo, University of California Berkeley*.
- Harmon, C. and Walker, I. (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *The American Economic Review*, 85(5):pp. 1278–1286.
- Harper, S., Lynch, J., Burris, S., and Davey Smith, G. (2007). Trends in the Black-White Life Expectancy Gap in the United States, 1983-2003. *JAMA*, 297(11):1224–1232.
- Harris, K. M. (2010). An Integrative Approach to Health. *Demography*, 47(1):1–22.
- Hayward, M., Pienta, A., and McLaughlin, D. (1997). Inequality in Men’s Mortality: The Socioeconomic Status Gradient and Geographic Context. *Journal of Health and Social Behavior*, 38(4):313–330.

- Howard, G., Anderson, R. T., Russell, G., Howard, V. J., and Burke, G. L. (2000). Race, Socioeconomic Status, and Cause-Specific Mortality. *Annals of Epidemiology*, 10(4):214–223.
- Hu, Y. and Goldman, N. (1990). Mortality Differentials by Marital Status: An International Comparison. *Demography*, 27(2):233–250.
- Hummer, R. A. (1996). Black-White Differences in Health and Mortality: A Review and Conceptual Model. *The Sociological Quarterly*, 37(1):105–125.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):pp. 467–475.
- Jacob, B. A. (2003). Are Idle Hands the Devil’s Workshop? Incapacitation, Concentration, and Juvenile Crime. *American Economic Review*, 93(5):1560–1577.
- Johnson, N. J., Backlund, E., Sorlie, P. D., and Loveless, C. A. (2000). Marital Status and Mortality: The National Longitudinal Mortality Study. *Annals of Epidemiology*, 10(4):224–238.
- Johnson, N. J., Sorlie, P. D., and Backlund, E. (1999). The Impact of Specific Occupation on Mortality in the U.S. National Longitudinal Mortality Study. *Demography*, 36(3):355–367.
- Kalwij, A. (2010). The impact of family policy expenditure on fertility in western Europe. *Demography*, 47:503–519.
- Keith, V. M. and Smith, D. P. (1988). The Current Differential in Black and White Life Expectancy. *Demography*, 25(4):625–632.
- Kimbrow, R., Bzostek, S., Goldman, N., and Rodríguez, G. (2008). Race, Ethnicity, and the Education Gradient in Health. *Health Affairs*, 27(2):361.

- Kitagawa, E. M. (1955). Components of a Difference Between Two Rates. *Journal of the American Statistical Association*, 50(272):1168–1194.
- Kowalski, A. E. (2009). Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care. Working Paper 15085, National Bureau of Economic Research.
- Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.
- Lefgren, L. and McIntyre, F. (2006). The Relationship between Women’s Education and Marriage Outcomes. *Journal of Labor Economics*, 24(4):pp. 787–830.
- Lillard, L. A. and Panis, C. W. A. (1996). Marital Status and Mortality: The Role of Health. *Demography*, 33(3):313–327.
- Lleras-Muney, A. (2005). The Relationship between Education and Adult Mortality in the United States. *The Review of Economic Studies*, 72(1):pp. 189–221.
- Machin, S., Marie, O., and Vujić, S. (2011). The Crime Reducing Effect of Education. *The Economic Journal*, 121(552):463–484.
- Mazumder, B. (2008). Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws. *Federal Reserve Bank of Chicago Economic Perspectives*, 32(2):2–16.
- McCrary, J. (2011). The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth. *American Economic Review*, 101(1):158–95.
- Meara, E. R., Richards, S., and Cutler, D. M. (2008). The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981-2000. *Health Aff*, 27(2):350–360.

- Mincer, J. (1963). Market Prices, Opportunity Costs, and Income Effects. In *Measurement in Economics: Studies in Mathematical Economics and Econometrics in memory of Yehuda Grunfeld*. Stanford University Press.
- Murray, J. E. (2000). Marital Protection and Marital Selection: Evidence from a Historical-Prospective Sample of American Men. *Demography*, 37(4):511–521.
- National Center for Health Statistics. Life Tables, 1987. *Vital statistics of the United States, 1987: Life tables (Vol. II, Section 6)*.
- National Center for Health Statistics. Life Tables, 1988. *Vital statistics of the United States, 1988: Life tables (Vol. II, Section 6)*.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.
- Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter. *The American Economic Review*, 96(1):pp. 152–175.
- Osili, U. O. and Long, B. T. (2008). Does Female Schooling Reduce Fertility? Evidence from Nigeria. *Journal of Development Economics*, 87(1):57 – 75.
- Painter, G. and Levine, D. I. (2004). Daddies, Devotion, and Dollars. *American Journal of Economics and Sociology*, 63(4):813–850.
- Pauly, M. V. and Herring, B. J. (2000). An efficient employer strategy for dealing with adverse selection in multiple-plan offerings: an MSA example. *Journal of Health Economics*, 19(4):513 – 528.
- Potter, L. B. (1991). Socioeconomic Determinants of White and Black Males' Life Expectancy Differentials, 1980. *Demography*, 28(2):303–321.

- Preston, S., Heuveline, P., and Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Wiley-Blackwell.
- Richardus, J. and Kunst, A. (2001). Black-White Differences in Infectious Disease Mortality in the United States. *American Journal of Public Health*, 91(8):1251.
- Rogers, R. G. (1992). Living and Dying in the U.S.A.: Sociodemographic Determinants of Death Among Blacks and Whites. *Demography*, 29(2):287–303.
- Rogot, E., Sorlie, P. D., and Johnson, N. J. (1992). Life Expectancy by Employment Status, Income, and Education in the National Longitudinal Mortality Study. *Public Health Reports (1974-)*, 107(4):457–461.
- Royer, H. (2004). What All Women (and Some Men) Want to Know: Does Maternal Age Affect Infant Health? Working paper no. 68, University of California, Berkeley, Center for Labor Economics.
- Sorlie, P. D., Backlund, E., and Keller, J. B. (1995). US Mortality by Economic, Demographic, and Social Characteristics: The National Longitudinal Mortality Study. *American Journal of Public Health*, 85(7):949–956.
- Train, K. (2009). *Discrete Choice Methods With Simulation*. Cambridge University Press.
- Waldron, I., Hughes, M. E., and Brooks, T. L. (1996). Marriage Protection and Marriage Selection—Prospective Evidence for Reciprocal Effects of Marital Status and Health. *Social Science & Medicine*, 43(1):113–123.
- Warner, D. and Hayward, M. (2006). Early-life Origins of the Race Gap in Men’s Mortality. *Journal of Health and Social Behavior*, 47(3):209–226.
- Williams, D. R. and Collins, C. (1995). US Socioeconomic and Racial Differences in Health: Patterns and Explanations. *Annual Review of Sociology*, 21(1):349–386.

- Williams, D. R. and Collins, C. (2001). Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. *Public Health Reports*, 116(5):404.
- Willis, R. J. (1973). A New Approach to the Economic Theory of Fertility Behavior. *Journal of Political Economy*, 81(2):S14–64.
- Winkleby, M. and Cubbin, C. (2003). Influence of Individual and Neighbourhood Socioeconomic Status on Mortality Among Black, Mexican-American, and White Women and Men in the United States. *British Medical Journal*, 57(6):444.