Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

1-1-2015

What Do Longitudinal Data on Millions of Hospital Visits Tell Us about the Value of Public Health Insurance as a Safety Net for the Young and Privately Insured?

Amanda E. Kowalski

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Kowalski, Amanda E., "What Do Longitudinal Data on Millions of Hospital Visits Tell Us about the Value of Public Health Insurance as a Safety Net for the Young and Privately Insured?" (2015). *Cowles Foundation Discussion Papers*. 2404.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2404

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

WHAT DO LONGITUDINAL DATA ON MILLIONS OF HOSPITAL VISITS TELL US ABOUT THE VALUE OF PUBLIC HEALTH INSURANCE AS A SAFETY NET FOR THE YOUNG AND PRIVATELY INSURED?

By

Amanda E. Kowalski

January 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1983



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

What Do Longitudinal Data on Millions of Hospital Visits Tell us About The Value of Public Health Insurance as a Safety Net for the Young and Privately Insured?*

Amanda E. Kowalski
Department of Economics, Yale University and NBER

January 16, 2015

Abstract

Young people with private health insurance sometimes transition to the public health insurance safety net after they get sick, but popular sources of cross-sectional data obscure how frequently these transitions occur. We use longitudinal data on almost all hospital visits in New York from 1995 to 2011. We show that young privately insured individuals with diagnoses that require more hospital visits in subsequent years are more likely to transition to public insurance. If we ignore the longitudinal transitions in our data, we obscure over 80% of the value of public health insurance to the young and privately insured.

*Kate Archibald, Norma Padron, and Samuel Moy provided excellent research assistance. Jason Abaluck, Joseph Altonji, Costas Meghir, Jonathan Skinner, Heidi Williams, and seminar participants at Yale and the AEA Meetings provided helpful comments. John Piddock and Larry Schoen provided invaluable assistance with the data. The National Science Foundation CAREER Award 1350132 and the Yale Department of Economics Program in Applied Policy provided research support. Contact information: amanda.kowalski@yale.edu, 202-670-7631.

1 Introduction

The current private health insurance system does not offer full long-term protection against financial risk, with most contracts lasting for only one year. If a young privately insured person gets a very severe illness in a given year, his private health insurance generally protects him against large financial losses in that year. However, he could lose private coverage in subsequent years for a variety of reasons. For example, he could face a lifetime limit on benefits, he could become so sick that he loses his job and his employer-sponsored coverage, or his premiums could increase so much that he can no longer afford to pay them. Several economists have acknowledged that the private health insurance system does not offer long-term financial protection for the nonelderly, and they have proposed approaches to address it through the design and regulation of private health insurance contracts (see Cochrane (1995) and Pauly et al. (1995)). In this paper, we consider public health insurance as an alternative approach to fill gaps in the current system. We focus on how public health insurance acts as a "safety net" that reduces financial risk for individuals who start out young and privately insured.

Previous literature has not quantified how many young privately insured individuals transition to public health insurance over periods of several years, and it has not quantified the value of public insurance for this population. Instead, the literature focuses on the value of the safety net to individuals who have already been "caught" by it (see Barcellos and Jacobson (2014), Khwaja (2010), Finkelstein and McKnight (2008) on Medicare; Engelhardt and Gruber (2011) on Medicare Part D; and recent work by Finkelstein et al. (2014) on Medicaid). These papers compare individuals with public insurance to those without public insurance. However, they do not consider that even individuals who are not covered by public insurance might value their potential access to it in the future. This limitation arises in part because the literature generally only uses cross-sectional data.

The main contribution of our paper is that we use longitudinal data on almost all hospital visits in the state of New York from 1995 to 2011 to inform the value of public health insurance as a safety net for the young and privately insured. Using our longitudinal data, we examine transitions in hospital spending and health insurance among young individuals with private health insurance. We show that individuals with diagnoses that require multiple future hospital visits have costs that are highly correlated over time, and they are more likely to gain public health insurance. This descriptive analysis provides motivation for the value of public health insurance to the young and privately insured. Then, using our data as the main input, we estimate the value of public health insurance to this population. We find that we miss over 80 percent of the value of insurance if we do not take longitudinal transitions into account. Even when longitudinal data are available but the panel length is short and the sample size is small, as in the widely-used Medical Expenditure Panel Survey (MEPS), standard calculations continue to miss most of the value of insurance. Intuitively, small datasets with short panel lengths offer limited information about observations in the tails of the distribution, which drive the value of insurance.

Our insights on the value of insurance complement those in other literatures that examine costs associated with expansions in public health insurance. In their seminal paper, Cutler and Gruber

(1996), estimate the extent to which expansions in public health insurance "crowd out" private health insurance at the time of public health insurance expansions. However, their data do not allow them to measure "crowd out" that occurs years, sometimes decades, later. Furthermore, their paper and many papers that build on it generally cast "crowd out" in a negative light. Similarly, a large theoretical and empirical literature shows that the public health insurance safety net could depress saving (see Hubbard et al. (1995), Golosov and Tsyvinski (2006) and De Nardi et al. (2010)). From an individual's perspective, these savings reductions may reflect the *value* of the potential "crowd out" of private health insurance that we estimate.

In Section 2, we present stylized facts from our data that motivate the value of public health insurance to the young and privately insured. In Section 3 we discuss the framework that has been used in the literature to estimate the value of health insurance, and we extend it to incorporate several periods. In Section 4, we discuss how we incorporate all of the individual-level longitudinal profiles observed in our data into our empirical framework. In Section 5, we present results from our framework, and we compare the value of health insurance obtained using longitudinal data to the value that would be obtained with cross-sectional data. We discuss the implications of our results and conclude in Section 6.

2 Stylized Facts

We present stylized facts that rely on the longitudinal nature of our data to motivate the value of public health insurance as a safety net for the young and privately insured. Our main data are restricted use data from the Statewide Planning and Research Cooperative System (SPARCS), which contain data on almost all inpatient discharges in the state of New York from 1995 to 2011. We supplement these data with data from other sources to construct a balanced panel designed to represent all individuals in the state of New York from 1995 to 2011. We focus on individuals who were ages 25-34 in 1995 to isolate young but working-age individuals. We further restrict our sample to individuals with private insurance in 1995; thus, our main "private-in-1995" sample consists of young individuals for whom public health insurance has value only as a potential safety net. In our initial analysis, we also exclude individuals who subsequently go uninsured because their exclusion simplifies the value of the public health insurance safety net and helps to build intuition. However, we include individuals who subsequently go uninsured in our preferred results. We provide further details on our data in Appendix A.

Using our data, we want to follow those individuals who are likely drive the value of the public health insurance safety net: individuals with persistent medical costs across multiple years. We perform a simple exercise to isolate those individuals. For individuals with at least one hospital visit in 1995, we determine the primary diagnosis for their most costly hospital visit. We then calculate the number of subsequent years for which visitors with that diagnosis in 1995 appear in the sample. We prefer a count of the number of subsequent years instead of the subsequent number of years with a given diagnosis in case some diagnoses manifest themselves as other diagnoses over time. We rank diagnoses according to this metric and classify diagnoses in the top 3% of this ranking

as "persistent DX" using 3 digit codes from the International Classification of Diseases (ICD-9). We then group individuals whose primary 1995 diagnosis is a "persistent DX" into the category "1995 Visit, Persistent DX." We group all other individuals with a 1995 hospital visit into the category "1995 Visit, Other DX."[1]

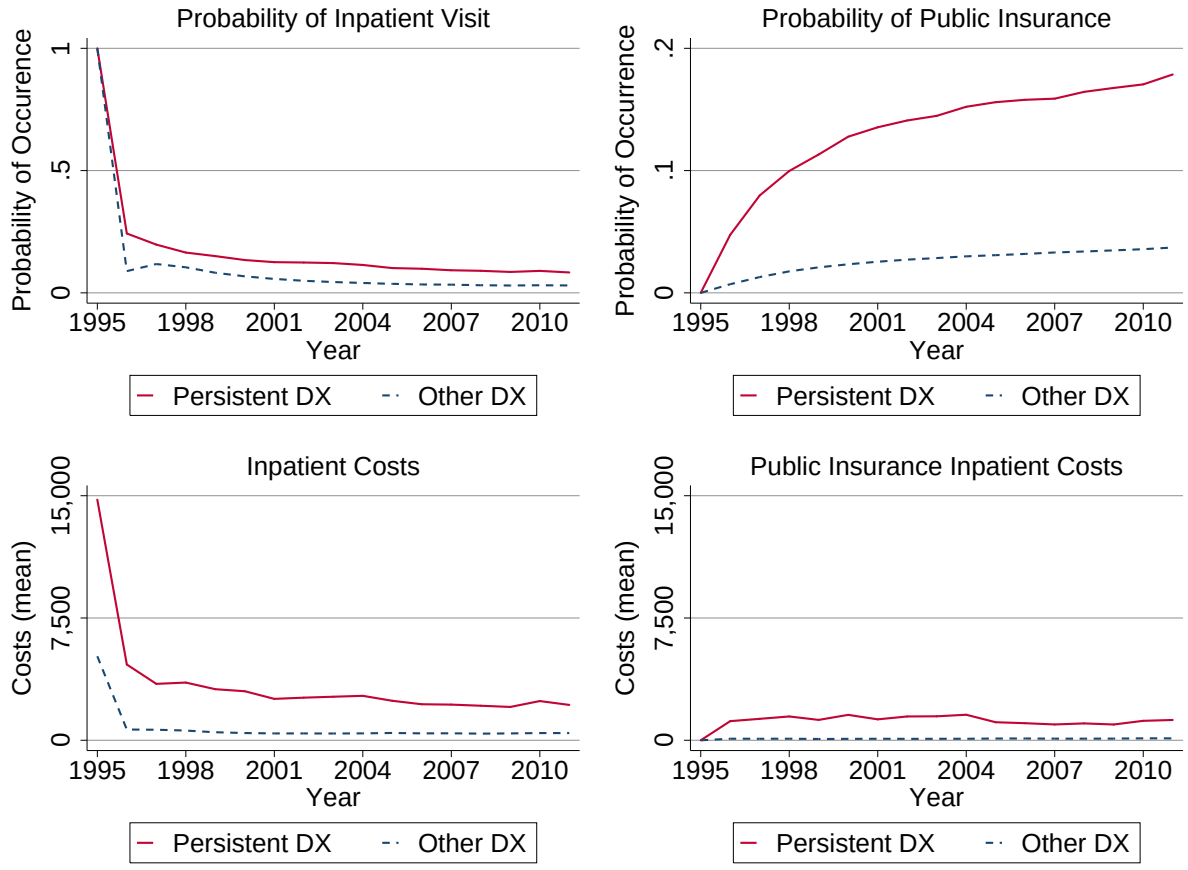
Table 1 lists the most common persistent diagnoses in our sample. The full set includes over 90 diagnoses, but the table depicts only those for which there are 30 or more individuals in our data with a 1995 diagnoses. Chronic kidney disease is the diagnosis for which we observe the highest rate of return to the hospital in subsequent years. In the 16 subsequent years we observe through 2011, among the individuals in our data with a primary diagnosis of chronic kidney disease in 1995, the average number of subsequent years with hospital visits through 2011 is 4.9. Other diagnoses, such as hereditary anemia and cystic fibrosis, also portend a high number of subsequent years of hospital visits. These visits do not all occur in the years immediately following 1995, so five years of panel data would not capture all of them. Figure 1 follows various outcomes for individuals with persistent diagnoses and all other diagnoses longitudinally over the entire length of our panel.

Table 1: Persistent DX

ICD-9 Diagnosis Code	Description	Average Subsequent Years with Hospital Visits through 2011
403	Chronic Kidney Disease (hypertensive renal disease)	4.9
282	Hereditary Anemia	4.7
277	Cystic Fibrosis	3.6
707	Ulcers	3.4
710	Lupus	3.4
340	Multiple Sclerosis	3.2
250	Diabetes	3.1
414	Heart Disease (Coronary Atherosclerosis)	3.1
595	Urinary Tract Infection	3.1
345	Epilepsy	3.0
295	Schizophrenic Disorders	2.9
555	Inflammation of Intestinal Tract	2.9
996	Implant and Graft Complications	2.9
571	Liver Disease	2.8
568	Peritoneal Adhesions	2.8
576	Disorders following Removal of Gallbladder	2.8
724	Spinal Pain (Spinal Stenosis)	2.8
569	Diseases of Anus, Rectum, and Intestinal Tract	2.7
723	Chronic Neck Pain	2.7
577	Diseases of the Pancreas	2.7
337	Peripheral Neuropathy	2.7
536	Diseases of Stomach (Achlorhydria)	2.7
349	Diseases of Central Nervous System	2.6
304	Drug Addiction	2.6

[1]There are 4,655 individuals in the "1995 Visit, Persistent DX" category (3.05% of hospital visitors in our private-in-1995 sample in 1995, and 0.28% of the entire private-in-1995 sample). There are 148,159 individuals in the "1995 Visit, Other DX" category (97.0% of hospital visitors in the private-in-1995 sample in 1995, and 8.77% of the entire private-in-1995 sample). There are 1,536,264 remaining individuals who do not have a visit in 1995 (91.0% of the entire sample).

Figure 1: Longitudinal Transitions for Individuals with 1995 Visit, Persistent DX vs. Other DX



Probability of Inpatient Visit The subfigure in the upper left of Figure 1 verifies that we have mechanically selected persistent diagnoses for which a large percentage of individuals show up in the hospital in subsequent years. Together, the persistent diagnosis and other diagnosis categories include all individuals in our sample who visit the hospital in 1995. Although individuals must have a hospital visit in 1995 to be classified into either diagnosis category, individuals with persistent diagnoses are much more likely to have hospital visits in subsequent years. Among individuals with persistent diagnoses, 24.3% have a subsequent visit the following year in 1996, 12.6% have a subsequent visit five years later in 2001, and 8.4% have a subsequent visit in 2011, the final year of our data. As shown in the dashed line, the rate of subsequent visits is lower for individuals with other diagnoses.[2]

Inpatient Costs Since our framework for valuing public health insurance will be based on costs and not visits, we verify that the visit patterns that we observe in each group translate

[2]The spike in the "1995 Visit, Other DX" category in 1997 is due to second pregnancies. This spike disappears when we exclude deliveries and other pregnancy-related diagnoses from the data.

into similar patterns for costs.[3] In the bottom left subfigure, we show mean inpatient costs after imposing a $30,000 maximum to annual costs.[4] As shown, the cost patterns do mirror the visit patterns. Cumulative mean costs are $58,656 for individuals with persistent diagnoses and $12,672 for individuals with other diagnoses.

Probability of Public Insurance All of the individuals in our sample begin with private health insurance in 1995, so the probability of having public health insurance must weakly increase relative to 1995 for all groups, as individuals switch to having public coverage. The top right subfigure of Figure 1 shows that individuals with persistent diagnoses are much more likely than individuals with other diagnoses to switch to public health insurance in subsequent years. Roughly 17.9% of individuals with persistent diagnoses in 1995 have public coverage by 2011, as compared to approximately 3.7% of individuals with other diagnoses.

Public Insurance Inpatient Costs In the final subfigure of Figure 1, having shown that individuals with persistent diagnoses have higher average costs and a higher probability of switching to public coverage than individuals with less persistent diagnoses, we show that these individuals also have higher average costs paid for by public insurance in each year. Across the entire time period, cumulative mean public insurance inpatient costs are $3,750 in the entire sample. Cumulative mean public costs are $20,098 for individuals with persistent diagnoses and $1,560 for individuals with other diagnoses. This subfigure is on the same scale as the subfigure beside it that depicts total inpatient costs, emphasizing that individuals who are privately insured in 1995 but appear at the hospital with a diagnosis that will result in persistent medical costs in future years end up having a large share of their future bills paid by public health insurance.

These stylized facts motivate the value of public health insurance as a safety net. Few other data sources could produce stylized facts like these because few other data sources track the medical spending and insurance coverage of nonelderly individuals over periods of greater than 10 years. Even those few sources that do follow the same individuals for more than 10 years do not capture the transitions from private to public coverage necessary to elucidate the longitudinal value of public health insurance to the privately insured. For example, claims data from firms often track individuals for many years, but these data only track individuals if they remain covered through the same firm. Thus, individual coverage transitions are unobservable in these sources of firm data, likely eliminating data on precisely those individuals with the highest realized valuations of the public health insurance safety net. Similarly, Medicare claims data follow individuals longitudinally for many years, but only when they are covered by public health insurance, leading crucial coverage transitions to be unobservable. The widely-used Medical Expenditure Panel Survey (MEPS) follows nonelderly individuals longitudinally, but only for 2.5 years. Using the framework for valuing health

[3]We observe hospital list prices rather than costs in our data, but we use the standard practice to deflate list prices into costs, as we discuss in Appendix A.

[4]We report average costs over all individuals in each cohort — even those who have died.

insurance that we develop below and our 17-year-long panel of data, we show that 2.5 year is not long enough to capture the vast majority of the longitudinal value of the public health insurance safety net to the young and privately insured.

3 Framework for Valuing Public Health Insurance as a Safety Net for the Privately Insured

To determine an individual's willingness to pay for health insurance, all papers in the literature use some variant of the following simple indifference condition, which gives expected utility in the presence of insurance on the left and expected utility in the counterfactual absence of insurance on the right:

$$\frac{1}{N}\sum_{i=1}^{N} u\left(\left[\sum_{t=1}^{T} I_t - c_{it}(M)\right] - \rho \middle| \alpha\right) = \frac{1}{N}\sum_{i=1}^{N} u\left(\sum_{t=1}^{T} I_t - c'_{it}(M) \middle| \alpha\right). \tag{1}$$

This equation determines ρ, the value that an individual is willing to pay to retain the world with health insurance. There are T time periods t and N individuals i, each of whom has an actual out of pocket cost profile of $\{c_{it}\}_{t=1}^{T}$ in the presence of health insurance and a counterfactual out of pocket cost profile $\{c'_{it}\}_{t=1}^{T}$ in the absence of health insurance. To allow for discounts to the insured and uninsured, costs are bounded above by M in each period. The utility-maximizing individual knows the distribution of longitudinal cost profiles with certainty; however, she does not know which realization will be hers (she does not know **which** $i \in \{1, \ldots, N\}$ she will be). The individual calculates her expected utility according to the utility function $u(x|\alpha)$, under the assumption that each realization i is equally likely, where α is a risk aversion parameter. All uncertainty in consumption occurs through uncertainty in out of pocket costs. The term I_t gives the individual's income in period t, and she knows her income profile $\{I_t\}_{t=1}^{T}$ with certainty. If utility takes on the constant absolute risk aversion (CARA) functional form, where $u(x|\alpha) = 1 - e^{-\alpha x}$, I_t drops out of the equation,[5] and a closed form solution exists for ρ as follows:

$$\rho = \frac{\ln\left(\dfrac{\sum_{i=1}^{N} e^{\alpha(\sum_{t=1}^{T} c'_{it}(M))}}{\sum_{i=1}^{N} e^{\alpha(\sum_{t=1}^{T} c_{it}(M))}}\right)}{\alpha} \tag{2}$$

As we demonstrate using data, for any nonzero level of risk aversion, this framework can yield drastically different values for the willingness pay to avoid the world without health insurance, ρ, depending on whether the actual cost profiles are obtained from cross-sectional or longitudinal

[5] Proof:

$$\frac{1}{N}\sum_{i=1}^{N}[1 - e^{-\alpha([\sum_{t=1}^{T} I_t - c_{it}(M)] - \rho)}] = \frac{1}{N}\sum_{i=1}^{N}[1 - e^{-\alpha[\sum_{t=1}^{T} I_t - c'_{it}(M)]}]$$

implies

$$e^{-\alpha(\sum_{t=1}^{T} I_t)} e^{\alpha\rho} \frac{1}{N}\sum_{i=1}^{N} e^{\alpha(\sum_{t=1}^{T} c_{it}(M))} = e^{-\alpha(\sum_{t=1}^{T} I_t)} \frac{1}{N}\sum_{i=1}^{N} e^{\alpha(\sum_{t=1}^{T} c'_{it}(M))}.$$

data. To perform the comparison, we compare three special cases of ρ, as shown in Table 2. We first characterize the risk-neutral value of insurance. Next, given nonzero risk aversion, we compare the cross-sectional value of insurance to the longitudinal value of insurance.

Table 2: Special Cases of The Value of Insurance, ρ

	$\boldsymbol{\lambda}$	$\boldsymbol{\phi}$	$\boldsymbol{\gamma}$
	Risk-neutral Value of Insurance	Cross-sectional Value of Insurance	Longitudinal Value of Insurance
$\boldsymbol{c_{it}}$, $\boldsymbol{c_{it}^{'}}$	Cross-sectional / Longitudinal	Cross-sectional	Longitudinal
α	0	α	α

Risk-neutral Value of Insurance, $\boldsymbol{\lambda}$ First, consider the special case in which agents are not risk averse. In this special case, as we show, it does not matter if the cost profiles are derived from cross-sectional or longitudinal data. All of the other parameters in Equation 2, M, T, and N can take on their full range of values for discounts, time periods, and sample sizes, respectively. We refer to the resulting value of ρ as λ. Taking the limit of Equation 2 as the risk aversion parameter α approaches zero, the willingness to pay for insurance is simply the difference in average costs between the world with insurance and the world without insurance:

$$\lambda = \lim_{\alpha \to 0} \rho = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} c_{it}^{'}(M)}{N} - \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} c_{it}(M)}{N}. \tag{3}$$

This special case makes intuitive sense. Individuals with no risk aversion only care about the average cost in the world with health insurance relative to the average cost in the world without health insurance. In the case where the world with health insurance is a world where all insurance is public health insurance, for which individuals pay no premium, and the world without health insurance is a world with the same cost distribution but no health insurance, λ is simply average spending covered by public health insurance. In this special case, agents value public health insurance simply because it is a transfer; not because it protects them from risk.

Cross-sectional Value of Insurance, $\boldsymbol{\phi}$ Next, consider the case where agents are risk averse, so $\alpha > 0$, and the other parameters M, T, and N take on their full range of values. The value of health insurance ρ in this special case ϕ includes a risk premium over and above the risk-neutral value of insurance λ. Under this scenario, we suppose only cross-sectional data are available. Thus, in each period $t \in \{1, \ldots, T\}$, an agent draws an index $i_t \in \{1, \ldots, N\}$ and is assigned the corresponding costs in that period (c_{it} in the actual world and $c_{it}^{'}$ in the counterfactual world), where each $i_t \in \{1, \ldots, N\}$ is equally likely for each period t. (In the longitudinal case that we consider next, an agent draws a single index $i \in \{1, \ldots, N\}$ and is assigned the corresponding cost profile that gives actual and counterfactual costs in all T periods — $\{c_{it}\}_{t=1}^{T}$ in the actual world and $\{c_{it}^{'}\}_{t=1}^{T}$

in the counterfactual world.) In this cross-sectional case, agents know the distribution of costs with and without insurance in each period, but they do not know the distribution of longitudinal cost profiles. Most studies stop here in calculating the value of insurance, using only one period of data ($T = t = 1$).

Longitudinal Value of Insurance, γ For the general case in which we have longitudinal data that allow costs to be correlated across multiple periods, we refer to the resulting value of ρ as γ. Fixing the distribution of costs in the worlds with and without insurance in each period, if costs are not correlated at all over time, then γ will be equal to ϕ. However, if costs are correlated over time, then γ will likely be much larger than ϕ. Intuitively, the risk that agents face is greater if cost shocks are correlated over multiple periods; therefore, willingness to pay for insurance will be greater if cost shocks are correlated over multiple periods.

The size of the longitudinal value of insurance γ relative to the cross-sectional value of insurance ϕ is the main empirical question that we seek to answer in this paper. Specifically, we are interested in how large the full value of insurance is when we take into account the correlation of shocks over time, relative to how large it is when we do not take into account the correlation of shocks over time. We parameterize this concept in the following ratio:

$$\frac{\gamma - \phi}{\gamma},$$

holding M, T, and N constant. This ratio gives the fraction of the value of insurance that is unaccounted for if we do not use longitudinal data.

The main contribution of this paper is in the use of the longitudinal data. However, the generalization of the simple framework to incorporate longitudinal data advances the literature in a subtle but critical way. Consider the related simple framework employed by Handel, Hendel, and Whinston (2013) who use longitudinal data to examine the tradeoff between adverse selection and reclassification risk. In their framework, assuming CARA utility, the following indifference condition determines the willingness to pay for insurance, μ:

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} u(I - c_{it}(M) - \mu|\alpha) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} u(I - c_{it}^{'}(M)|\alpha). \tag{4}$$

Relative to the framework that we use in Equation 1, utility is additively separable across time (both summation terms are outside of the utility function), and income I does not vary with time. We demonstrate that in our data used as repeated cross-sections, for a plausible range of the risk aversion parameter α, the cross-sectional value of insurance from our framework ϕ is almost exactly the same as the alternative cross-sectional value of insurance μ given by the alternative framework in Equation 4. (They are not exactly the same because our framework is a nonlinear utility function of the expectation over time, and the alternative framework is the expectation over time of a nonlinear utility function.) However, with longitudinal data, both frameworks give very different values because the framework given by Equation 4 does not allow risk to be correlated over

time. In each period, under Equation 4, agents effectively take an independent draw from the cost distribution in that period, such that longitudinal data yields the same value of μ as repeated cross-sectional data. (To see this, note that in this framework, taking expected utility across time and then across individuals is mathematically equivalent to taking expected utility across individuals and then across time). Thus, it is crucial to our analysis that we use a framework such as the framework given by Equation 1, which allows longitudinal data to yield different estimates from cross-sectional data.

4 Empirical Implementation of Framework for Valuing Public Health Insurance as a Safety Net for the Privately Insured

We develop an empirical implementation of the framework introduced in Equation 1 to estimate the value of public health insurance using our data. Table 2 shows the data and parameters that we need to solve for each of the special cases of the value of health insurance ρ via Equation 2. We begin by specifying the actual out of pocket cost profile $\{k_{it}\}_{t=1}^{T}$ for every individual i in our data. To do so, we first obtain to the total (beneficiary plus insurer) cost profile $\{c_{it}\}_{t=1}^{T}$ for each individual i from the data, with corresponding coverage information. In each period t, an individual i can be privately insured, which we denote with the dummy variable r_{it}; publicly insured by the "safety net," which we denote with the dummy variable s_{it}; or uninsured, which occurs when $(1 - r_{it} - s_{it}) = 1$. Because our data contain no information on out of pocket costs, we translate total costs c_{it} into out of pocket costs k_{it} via a simple formula:

$$c_{it} = r_{it} * \pi_t + (1 - r_{it} - s_{it}) * \min(k_{it}, M) \tag{5}$$

where

$$\pi_t = \frac{\sum_{i=1}^{N} r_{it} * \min(k_{it}, M)}{\sum_{i=1}^{N} r_{it}}.$$

As shown by the formula in Equation 5, we assume that out of pocket costs are equal to zero for publicly insured individuals; costs are equal to the premium, π_t, for privately insured individuals; and as shown, the premium is equal to the average cost of the privately insured after discounts. For uninsured individuals, costs are equal to the minimum of the total cost k_{it} and the maximum cost before discounts M. In our main sample, there are no uninsured individuals, so Equation 5 simplifies to $c_{it} = r_{it} * \pi_t$, but we retain the full notation for use with data that include uninsured individuals.

For every individual i in our data, we also need to specify the counterfactual out of pocket cost profile $\{c'_{it}\}_{t=1}^{T}$ that would occur were the public health insurance safety net dismantled. It would not be realistic to assume that all individuals with public health insurance would go uninsured, especially since we focus on individuals who begin with private health insurance in 1995. It is likely that only some of them would go uninsured, while others would gain private health insurance. To consider this more realistic case, which does not artificially inflate the value of public health insurance, we

use a logit model to predict which individuals would go uninsured and which individuals would obtain private health insurance upon the dissolution of the safety net. The logit model gives us a probability of private health insurance $r_{it}^{'}$ for each person i in each period t.[6] Accordingly, we define the cost profile for individual i in the world without the public health insurance safety net, by expressing $c_{it}^{'}$ as follows:

$$c_{it}^{'} = r_{it}^{'} * \pi_{t}^{'} + (1 - r_{it}^{'}) * \min(k_{it}, M) \tag{6}$$

where

$$\pi_{t}^{'} = \frac{\sum_{i=1}^{N} r_{it}^{'} * \min(k_{it}, M)}{\sum_{i=1}^{N} r_{it}^{'}}.$$

Private health insurance premiums in this counterfactual world $\pi_{t}^{'}$ reflect the average costs of individuals enrolled. We assume no moral hazard, so that costs before discounts, k_{it}, are the same in both the world with public health insurance and the world without. This assumption is motivated by data limitations, but it is likely to be a reasonable assumption for inpatient care.

Next, we derive "cross-sectional" cost profiles $\{c_{it}\}_{t=1}^{T}$ and $\{c_{it}^{'}\}_{t=1}^{T}$ from our longitudinal data. In each period t, we randomly draw costs c_{it} and $c_{it}^{'}$ (with their associated insurance types) from the observed and counterfactual distributions, respectively. This procedure removes all longitudinal elements from each individual's cost profile in the worlds with and without public health insurance. Importantly, it also ensures that average costs paid by public health insurance are exactly the same in the cross-sectional and longitudinal profiles such that the risk neutral value of insurance is the same for both profiles.

One additional consideration is whether individuals have private information about the set of cost profiles from which they will draw. In our baseline scenario, since we have already restricted our sample to individuals who are young and privately insured, we assume that individuals have no private information that would further narrow the distribution of cost profiles that they expect. Each individual knows the full distribution of cost profiles, and she believes at the time that the value of insurance is calculated that she is equally likely to realize any of them. To incorporate private information, we can allow individuals to consider themselves a part of a smaller group G_j within the population N, of size N_j, and they expect only the cost profiles within that group. For example, individuals can expect only the cost profiles of women. When calculating the value of insurance, we obtain a separate willingness to pay for health insurance ρ_j for each group, according the following modified indifference condition:

$$\frac{1}{N_j} \sum_{i \in G_j} u\left(\left[\sum_{t=1}^{T} -c_{it}(M)\right] - \rho_j \middle| \alpha\right) = \frac{1}{N_j} \sum_{i \in G_j} u\left(\sum_{t=1}^{T} -c_{it}^{'}(M) \middle| \alpha\right). \tag{7}$$

In this scenario, we still assume that the premiums are calculated based on the full population

[6] Our model makes predictions based on the age, sex, race, ethnicity, and county of individuals who appear in the SPARCS data. In periods in which an individual does not appear in the data, we apply the probabilities from the most recent year in which she did appear.

— private information affects the willingness to pay for insurance but not its actual cost.[7] Private information is valuable insofar as it predicts cost. We consider the robustness of our results to an extreme form of private information - private information that the individual will have a hospital visit in the first year of the sample that will require a great deal of subsequent treatment, putting the individual into the "1995 Visit, Persistent DX" category considered in Section 2. We show that even this extreme form of private information does not alter our main results. Therefore, in our baseline scenario, facing the common issue that it is difficult to know what private information individuals have (see Abaluck and Gruber (2011) and Handel (2013)), we assume that individuals have no private information but that they know the full distribution.

We also assume that individuals know that they could die during the period. They know that if they die, they will not incur any hospital costs and that they will not pay any more premiums. By 2011, approximately 2.5% of individuals in our sample are deceased. In practice, this is a small enough share of the sample that alternative treatments of death have little impact on the results.

Finally, to estimate the three special cases of the value of public health insurance ρ described in Table 2, following Equation 1, we need to specify values of the parameters for the risk aversion parameter, α, the number of time periods T, maximum costs before discounts M, and the number of individuals in the pool N. For each of these parameters, we report results for a wide range of values, rather than relying on a single calibrated value. We are particularly interested in the robustness of $(\gamma - \phi)/\gamma$, the value of public insurance accounting for the correlation of health costs over time, relative to the value of public insurance without accounting for the correlation of health costs over time.

5 Results and Robustness

We begin by calculating the risk-neutral value of insurance λ, given baseline parameter values: maximum annual costs $M = \$30,000$; the number of years $T = 17$ (the full number of years in our sample), and $N = 1,689,078$ (the full number of individuals in the our sample). The risk-neutral value of insurance, which sets $\alpha = 0$ following Table 2, is \$2,633, which is equal to the average costs paid by public insurance over the time period, as discussed in Section 2.

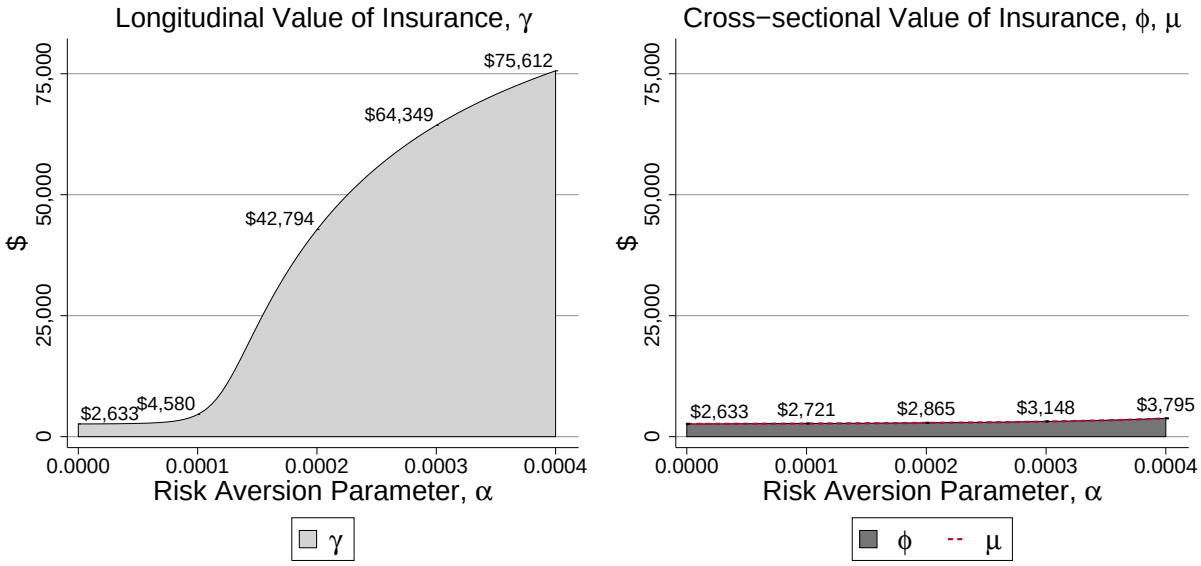
The longitudinal and cross-sectional values of insurance include this amount plus a risk premium, which depends on the risk aversion parameter α. We specify a baseline value of $\alpha = 0.0002$, and we examine robustness from 0 (the risk neutral case) to 0.0004, which is the median risk parameter estimated by Handel, Hendel, and Whinston (2013).[8] As shown in Figure 2, at $\alpha = 0$, the longitudinal value of insurance is equal to the risk-neutral value of insurance, average public

[7] Changing the premiums such that they only reflect individuals in each private information group would not be realistic because it would require individuals in each group to pool only with each other. Community rating laws in New York state would have hindered such pooling during our period of interest.

[8] Since Handel, Hendel, and Whinston (2013) focus on annual costs and utility, rather than combined costs and utility over a 17 year period, the costs in that paper are much smaller than the cumulative costs that we consider. The appropriate magnitudes of α are inversely related to the magnitudes of the losses in question. Since we are dealing with very large losses relative to Handel, Hendel, and Whinston (2013), 0.0004 should be a reasonable upper bound.

costs. As risk aversion increases, the longitudinal value of insurance increases smoothly from $4,580 at $\alpha = 0.0001$ to $42,794 at our baseline value $\alpha = 0.0002$ to $64,349 at $\alpha = 0.0003$ to $75,612 at the maximum that we consider $\alpha = 0.0004$. At our baseline value, the representative individual is willing to pay $42,794 to retain the public health insurance safety net before uncertainty is realized.

Figure 2: Longitudinal vs. Cross-sectional Values of Insurance



To give more context to our baseline estimate, we compare it to the cross-sectional value of insurance ϕ obtained with the same baseline parameter values. Figure 4 shows how the cross-sectional value of insurance ϕ varies with the risk aversion parameter. It varies from $2,633 when $\alpha = 0$ (the risk neutral case) to $2,865 at our baseline value of $\alpha = 0.0002$ to $3,795 at our maximum value of $\alpha = 0.0004$. For comparison, we also report the cross-sectional value of insurance μ obtained via Equation 4. As shown, the values are almost identical over the entire range with $\mu = \$2,847$ at our baseline value, only $18 less than ϕ. However, Equation 4 yields the same exact value with longitudinal data as it does with cross-sectional data. In contrast, in our framework, given by Equation 1, the longitudinal value of insurance $\gamma = \$42,794$ is much larger than the cross-sectional value of insurance $\phi = \$2,995$. At our baseline value of α, the fraction of the longitudinal value of health insurance that is not captured by the cross-sectional value of health insurance, $(\gamma - \phi)/\gamma$, is 93%.

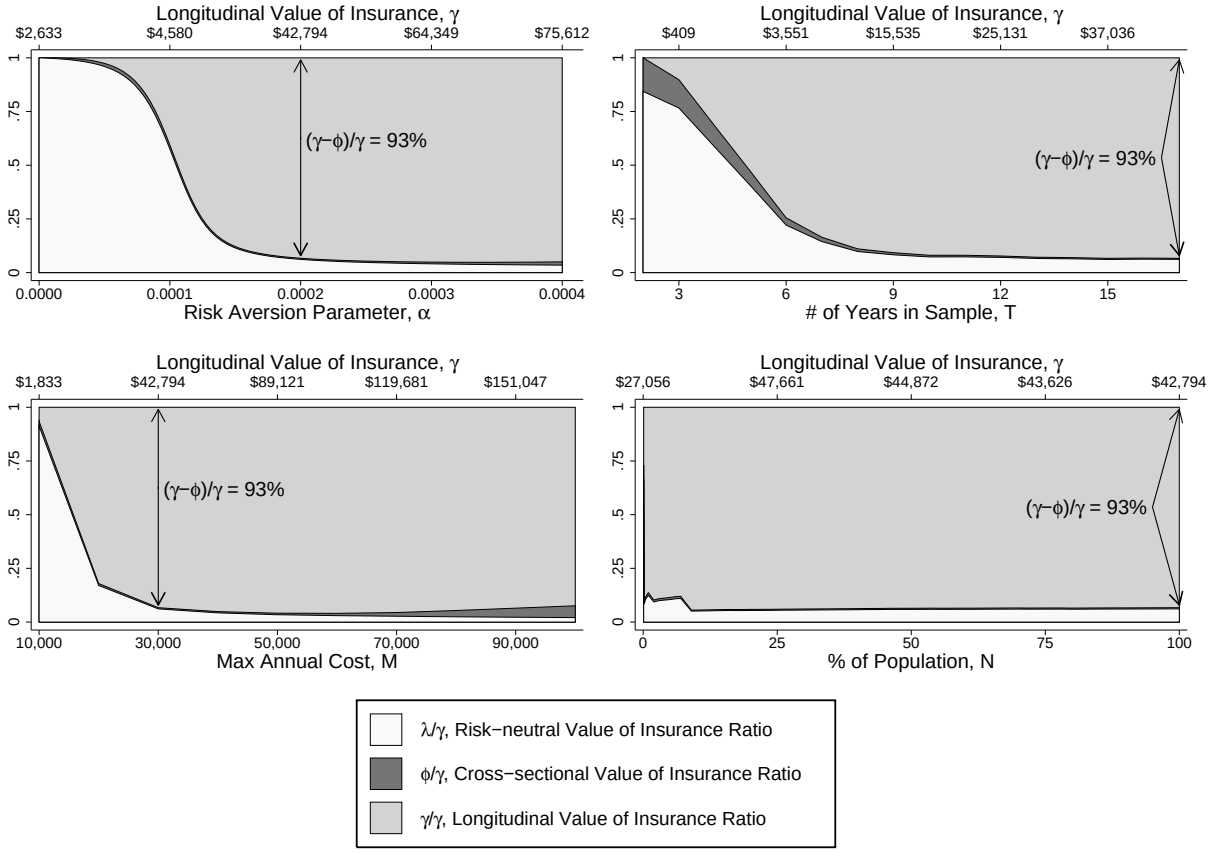
Although we have shown that the levels of the cross-sectional and longitudinal value of health insurance vary considerably with the risk aversion parameter α, the fraction of the longitudinal value of health insurance that is not captured by the cross-sectional value of health insurance, $(\gamma - \phi)/\gamma$, is much more robust. We focus on this ratio as our main result, and we gauge its robustness to variation in all aspects of the model. We consider robustness of this ratio to the risk aversion parameter α, to maximum annual cost M, to the number of years T, and to the sample size N. We then further consider robustness to private information and to the inclusion of individuals who go

uninsured.

Robustness to Risk Aversion Parameter, α In each subfigure of Figure 3, we report the risk neutral value of insurance λ as a fraction of the longitudinal value of insurance γ in the lightest shade, which is superimposed on the cross-sectional value of insurance ϕ as a fraction of the longitudinal value of insurance γ in the darkest shade. The remaining shade at the top reflects the portion of γ that is not captured in the other two parameters. Along the bottom axis, we vary the risk aversion parameter α, and we report the corresponding value of the longitudinal value of insurance γ along the top axis. We emphasize our main ratio of interest at our baseline parameter values $(\gamma - \phi)/\gamma$ by reporting the value with overlaid text.

As shown in the top left subfigure of Figure 3, even though willingness to pay for the public health insurance safety net continues to grow in absolute terms as the risk aversion parameter α increases, as soon as α reaches the middle of our range, the portion of the willingness to pay due to longitudinal risk, $(\gamma - \phi)/\gamma$, levels out at around 93% of the total willingness to pay. Therefore, for a very large range of α, willingness to pay for public insurance that only considers cross-sectional data severely understates willingness to pay for public insurance.

Figure 3: Robustness to α, M, T, and N



Robustness to Maximum Annual Costs, M Relative to α, which governs risk aversion, we expect the other three parameters, M, T, and N to be ancillary parameters. Nonetheless, we examine whether our results are sensitive to them. The bottom left subfigure 3 shows robustness to maximum costs after discounts, M. We vary M from \$10,000 through our baseline of \$30,000 to \$100,000. At the smallest values of M, the longitudinal value of insurance ϕ and the cross-sectional value of insurance γ are only a little larger than the risk neutral value of insurance λ. This occurs because in the absence of extreme values, risk decreases, as does willingness to pay for insurance, which reflects expected public cost more than it reflects a risk premium. However, starting at about $M = 30,000$, the proportion of willingness to pay due to risk starts to be fairly robust, motivating our choice of \$30,000 as our baseline value of M.

It is also interesting to note that at high values of M, the cross-sectional value of insurance reflected in ϕ/γ becomes larger relative to the risk neutral value of insurance, reflected in λ/γ because costs in each period have higher variability, even though their mean does not change much. However, ϕ/γ and λ/γ are still a small fraction of γ, suggesting that empirically, the longitudinal value of health insurance γ stems from people with moderate costs who visit the hospital regularly, rather than by people with high costs and infrequent visits.

Robustness to Number of Years, T In the upper right subfigure of Figure 3, we examine robustness to T, the number of years of data used, starting with 1995, from two years to the full 17 available years. We do not show robustness to one year of data because everyone in our sample knows that he will have private health insurance in 1995, so the value of public health insurance in that year is zero. As shown, after about 8 years, our main ratio of interest $(\gamma - \phi)/\gamma$ is extremely robust to number of years in the sample. Therefore, we use the full 17 years as our baseline value of T so that we use as much data as possible.

However, the results are not very robust to longitudinal data that uses fewer than eight years. Therefore, estimates of willingness to pay for insurance that use shorter periods of longitudinal data, such as the MEPS, which follows respondents for 2.5 years, will severely understate willingness to pay for the public health insurance safety net. Because shorter periods underestimate the importance of longitudinal risk, the impact of cross-sectional risk on the willingness to pay for public health insurance becomes more evident in these short periods. In results not shown, we examine robustness to different specifications of T, starting in years later than 1995, and the same general patterns hold.

Robustness to Sample Size, N Finally, in the bottom right subfigure of Figure 3, we examine the robustness of our results to sample size. In general, willingness to pay is quite robust to sample size, so we use our full sample of 1,689,078 as our baseline sample size to avoid throwing away information. As shown, a 10% sample is sufficient to get a very accurate distribution, and even a 1% is informative. However, sample sizes under 1% reflect drastically lower risk components of willingness to pay (likely due to the omission of tails of the distribution). As mentioned earlier, if we restrict the MEPS in a similar manner to our sample, we are left with a sample size that is approximately 0.3% the size of our data. Results at that sample size give highly variable estimates

of the longitudinal value of health insurance — a sample of 100 random draws of 0.3% of our data (about the size of the MEPS) yields a range of values of $(\gamma - \phi)/\gamma$ from 8.2% to 95.7%. It should also be noted that these results are at the MEPS sample size, assuming that the MEPS has 17 years of longitudinal data. As discussed above, limiting T to the actual 2.5 years of data in the MEPS obscures even more of the willingness to pay for public insurance.

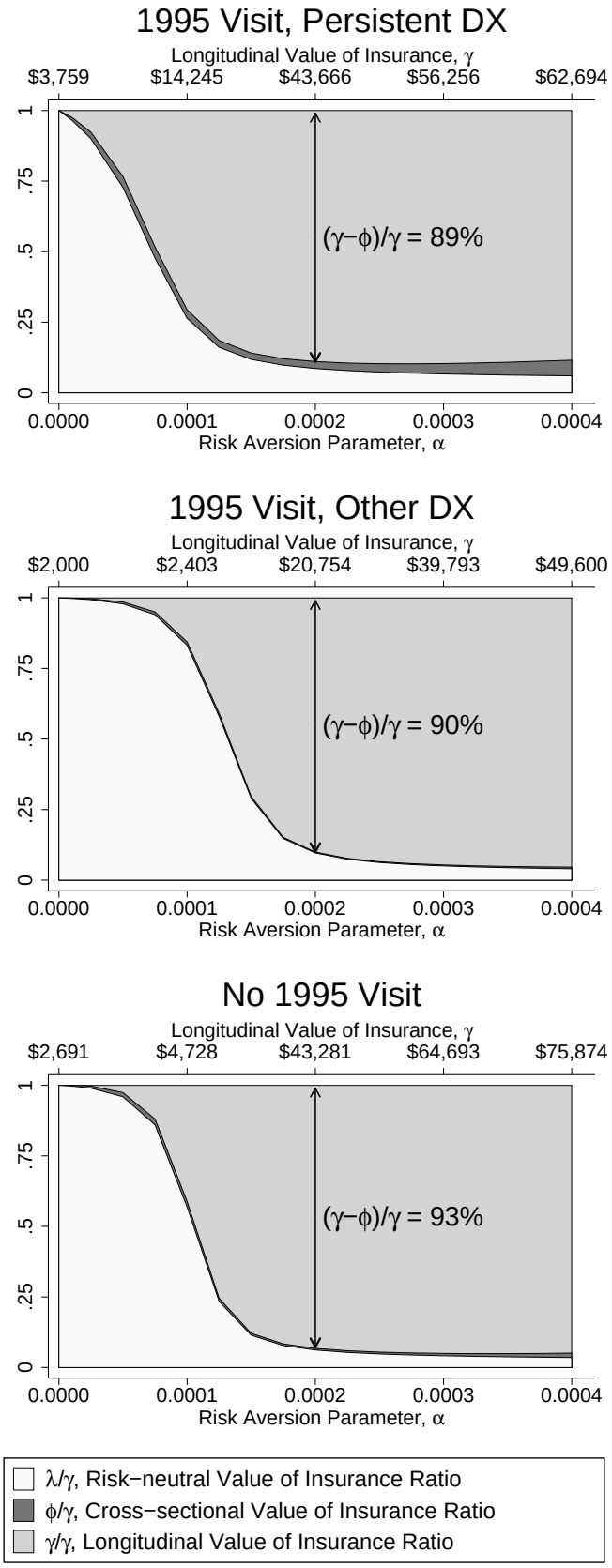
Robustness to Private Information on Persistent Diagnoses Our main specification assumes that everyone believes that she could have the same realized costs as anyone in the sample. In reality, we might think that individuals have private information about the cross-sectional and longitudinal distributions from which they will draw. To evaluate whether this private information has a meaningful impact on willingness to pay for the public health insurance safety net, we return to our diagnosis intensity cohorts (from Section 2) and consider the extreme case where we assume that the people who will enter the hospital in 1995 with the most persistent diagnoses know that they will enter the hospital with the most persistent diagnoses, forming a group G_j for calculation of the value of health insurance via Equation 7. We divide the full sample into two other groups: people who know that they will have other inpatient diagnoses in 1995 and people who know that they will not visit the hospital in 1995. Mechanically, we evaluate willingness to pay among these subgroups, while retaining the premiums from the overall sample.

Figure 4 shows how individuals value public insurance, knowing their diagnosis group. The top panel shows the robustness of willingness to pay to α for individuals who know that they will be in the "1995 Visit, Persistent DX" group, the middle panel shows the same figure for those who know that they will be in the "1995 Visit, Other DX" group, and the bottom panel shows the same figure for those who know that they will be in the "No 1995 Visit" group.

Unsurprisingly, the most striking difference across the persistence categories is the longitudinal value of public health insurance γ shown along the top axis. The longitudinal value of public health insurance γ is larger for those individuals with persistent diagnoses than it is for those individuals in the "1995 Visit, Other DX" category. The "No 1995 Visit" category, where there is no diagnosis in 1995, serves as a comparator category. It should closely resemble our main specification, as most individuals (91.0%) do not visit the hospital in 1995, and thus they have no private information. Indeed, we find nearly identical results in that group as we do in our main specification.

At low levels of the risk aversion parameter α, the persistent diagnosis group is the group with the largest longitudinal value of insurance γ, as shown along the top axis of each subfigure. However, a high levels of the risk aversion parameter α, "No 1995 Visit" group has the largest longitudinal value of insurance. To understand what drives this change, we note that the risk neutral value of insurance λ (the leftmost value of γ shown) is only slightly higher for the individuals with persistent diagnoses than it is for individual with no 1995 visit. As discussed above, the risk neutral value of insurance is simply the average costs ever paid by public health insurance. (Those costs are lowest in the "1995 Visit, Other DX" group, likely because of reversion to the mean after 1995 costs are paid by private health insurance). However, the *dispersion* of costs ever paid by public health insurance is larger among the "No 1995 Visit" group than it is in the persistent diagnosis group. Therefore, at

Figure 4: Robustness to Private Information on Persistent Diagnoses



higher levels of risk aversion, the group with private information about their persistent diagnoses has a lower valuation of public insurance than individuals with no private information.

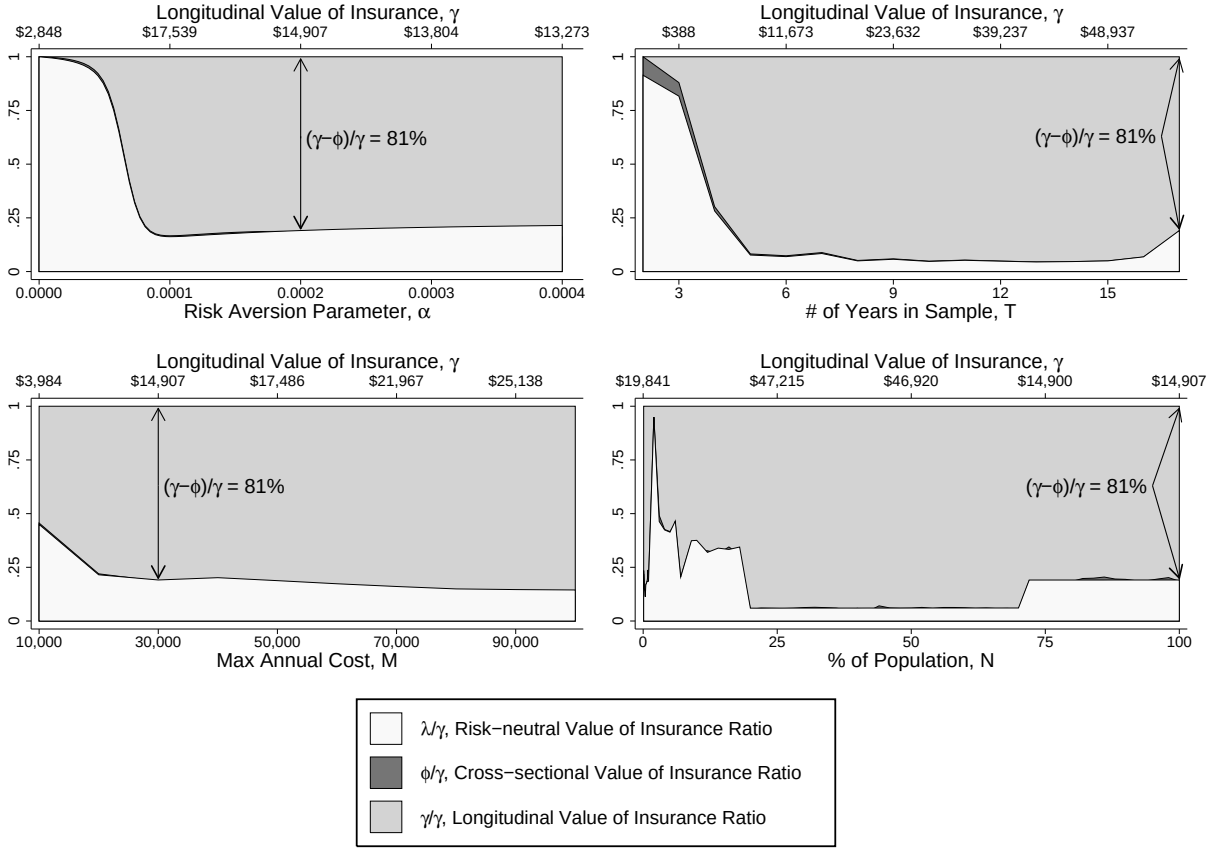
Even though the levels of the longitudinal value of health insurance γ differ across groups with differing amounts of private information, our general finding that cross sectional-data obscures a substantial fraction of the value of insurance holds. The fraction of longitudinal value of public health insurance that is obscured by cross-sectional data $(\gamma - \phi)/\gamma$ is 89%, among individuals with persistent diagnoses in 1995. Individuals with other inpatient diagnoses in 1995 have a ratio of 90%. We find quantitatively similar results if we use 2%, 1%, and 0.1% thresholds in lieu of the 3% threshold that we use to determine which diagnoses are persistent. Accordingly, we conclude that it is not constructive to explore other heterogeneity in private information, as persistence in diagnoses is an extreme form of private information, and it does not alter our main finding.

Robustness to Including Uninsured We began by excluding individuals who ever go uninsured from our private in 1995 sample because they make the willingness to pay for the public health insurance safety net less straightforward — individuals are less willing to pay for a safety net that has holes. Now that we have shown how willingness to pay varies when the safety net does not have holes, we add the uninsured back into the analysis, aiming to give a more realistic picture of the willingness to pay for the public health insurance safety net in the current environment. This larger sample includes an additional 134,750 individuals who are uninsured for some point during the period from 1996-2011, for a total balanced cohort population of 1,823,828

Figure 5 shows the same analysis as Figure 3, but it includes individuals who ever go uninsured. As shown in the numbers at the top of each subfigure, the longitudinal value of public health insurance is lower in dollar terms (as it should be) when we allow that public health insurance leaves some individuals uninsured. At our baseline parameter values, γ is \$42,794 when we do not include the uninsured, and it is \$14,907 when we do. However, the fraction of the longitudinal value of the public health insurance safety net that is not captured with cross-sectional data is only slightly smaller, and it is still very robust across a wide range of parameter values. We find that at $\alpha = 0.0002$, $(\gamma - \phi)/\gamma$ is equal to 81%, which is about 10% lower than the result using our primary "private-in-1995 sample."

The patterns evident in the subfigures that include the uninsured population are not quite as smooth as those for our main specification. The main phenomenon that drives the irregular patterns is that the longitudinal value of insurance γ and the cross-sectional value of insurance ϕ do not always increase monotonically with the risk aversion parameter α as they do in our primary sample. One manifestation of this non-monotonicity is that the longitudinal and cross-sectional values of insurance are not necessarily greater than the risk neutral value of insurance λ (though they are never significantly less than λ in the relevant range of α). To understand this phenomenon, it is useful to consider the limit of the value of insurance ρ, expressed in Equation 2, as the risk aversion parameter α approaches infinity. The limiting value is equal to the difference between the maximum cost in the counterfactual world and the maximum cost in the actual world:

Figure 5: Robustness to Including Uninsured



$$\lim_{\alpha \to \infty} \rho = \max_{i \in \{1,\ldots,N\}} \left[\sum_{t=1}^{T} c_{it}^{'}(M_t) \right] - \max_{i \in \{1,\ldots,N\}} \left[\sum_{t=1}^{T} c_{it}(M_t) \right] \tag{8}$$

Equation 8 demonstrates that it is theoretically possible for the longitudinal value of insurance ρ to be equal to zero even if the risk neutral value of insurance λ is positive. Intuitively, this is because the world with insurance leaves some individuals uninsured, reducing the willingness to pay for public insurance. Before any uncertainty is realized, the infinitely risk averse individual sees the tradeoff between the actual world and the counterfactual world pessimistically, and so he imagines the downside of being the person with the most costly health profile in each world, and considers the tradeoff between the two worlds under the assumption that he draws the short straw in either scenario. Importantly, willingness to pay in the case of extreme risk aversion is not the result of aversion to experiencing the biggest change in cost between the actual and counterfactual worlds based on a single health and insurance profile, but rather is driven by a risk-averse assessment of the set of cost profiles that exist in each world. If the individual with the highest costs will have the same costs in both worlds, then the infinitely risk averse individual has no willingness to pay for the safety net.

This intuition is important to keep in mind when considering the longitudinal value of insurance γ as it varies with the risk aversion parameter in the top left subfigure of Figure 5. If the individual with the highest cumulative costs by the end of the sample is always privately insured, then willingness to pay for public health insurance approaches zero as the risk aversion parameter α approaches infinity. Similarly, if the individual with the highest cumulative costs by the end of the sample is uninsured in some periods regardless of the presence of the safety net, then willingness to pay for public health insurance also approaches zero as the risk aversion parameter α approaches infinity. Thus, willingness to pay when taking risk aversion into account can actually be less than the risk neutral value of insurance (γ or ϕ can be less than λ).

As shown in the other subfigures of Figure 5, the fraction of longitudinal value of health insurance that is obscured by cross-sectional data $(\gamma - \phi)/\gamma$ does not always vary smoothly. This is simply another manifestation of the intuition that we have developed by taking the limit of the value of public health insurance as the risk aversion parameter α increases. Even at our baseline value of the risk aversion paramter α, the value of public health insurance depends on the tails of the distribution, and the tails change as we vary the sample and the maximum annual cost. The discrete jumps evident in the lower right subfigure of Figure 5 occur because as we increase the sample size N, most of the individuals we add fall in the interior of the distribution, having a smooth impact on the valuation of insurance. However, when we add an individual to the sample that changes the tail of the distribution, the value of insurance changes discretely.

Because of the nuance involved in understanding the framework when we allow individuals to go uninsured in the world with the public health insurance safety net, we introduced the uninsured after building up intuition in the simpler case. However, consider the results that include the uninsured to be our primary results. At the baseline values of our parameters, in the sample that includes the uninsured, the longitudinal value of insurance of \$14,907 is much lower than the longitudinal value of insurance of \$42,794 obtained in the restricted sample. However, whether we include the uninsured or not, we still miss a substantial fraction of the value of the public health insurance safety net if we focus on cross-sectional, rather than longitudinal data. Across a wide range of parameters, using cross-sectional data in lieu of longitudinal data obscures over 80% of the value of public health insurance to the young and privately insured.

6 Conclusion

The expansion of Medicaid coverage through the Affordable Care Act will expand the public health insurance safety net in many states. This expansion in public health insurance will not only affect individuals who are currently uninsured, but also it will serve as a safety net for individuals who currently have private coverage. Past estimates have fallen short in capturing the value of public health insurance as a safety net because they have used cross-sectional data to obtain their estimates. Such data sets fail to capture the fact that some individuals' medical costs are highly correlated over time, thereby increasing financial risk and the value of public health insurance that mitigates such risk. Furthermore, the size of most cross-sectional datasets used in the literature is small, offering

a limited picture of events that occur in the tail of the distribution and drive risk.

We employ a longitudinal data set that includes nearly all hospital visits in the state of New York from 1995 to 2011 to estimate the value that the young and privately insured derive from public health insurance. We consider individuals who are privately insured in 1995 but whose health insurance coverage may change over time. In 1995, public health insurance serves only as a potential safety net that mitigates future financial risk. We show that the individuals who have diagnoses in 1995 that are likely to require hospital care in future years are more likely to transition to public health insurance. We find that using cross-sectional data in lieu of longitudinal data obscures over 80% of the value of public health insurance to the young and privately insured.

This paper demonstrates that it is important to collect and analyze longitudinal data to understand the risk-protective value of health insurance. Our empirical analysis shows that patterns visible in longitudinal data that are not visible in cross-sectional data are economically important. Given this finding, models of the risk-protective value of health insurance should allow for correlations in coverage and costs across time. Beyond elucidating the value of the public health insurance safety net, this paper helps to explain why individuals value regulations that guarantee renewability of private health insurance regardless of pre-existing conditions — the empirical results show that the longitudinal value of health insurance coverage is much greater than the sum of its cross-sectional parts.

References

Jason Abaluck and Jonathan Gruber. Heterogeneity in choice inconsistencies among the elderly: evidence from prescription drug plan choice. *The American Economic Review*, 101(3):377–381, 2011.

Silvia Helena Barcellos and Mireille Jacobson. The effects of medicare on medical expenditure risk and financial strain. Technical report, National Bureau of Economic Research, 2014.

John H. Cochrane. Time-consistent health insurance. *Journal of Political Economy*, 103(3):pp. 445–473, 1995. ISSN 00223808. URL `http://www.jstor.org/stable/2138695`.

David Cutler and Jonathan Gruber. Does public insurance crowd out private insurance? *Quarterly Journal of Economics*, 111(2):391–430, 1996.

Mariacristina De Nardi, Eric French, and John B. Jones. Why do the elderly save? the role of medical expenses. *The Journal of Political Economy*, 118(1):39–75, 2010.

Gary V. Engelhardt and Jonathan Gruber. Medicare part d and the financial protection of the elderly. *American Economic Journal: Economic Policy*, 3(4):77–102, 2011.

Amy Finkelstein and Robin McKnight. What did medicare do? the initial impact of medicare on mortality and out of pocket medical spending. *Journal of Public Economics*, 92(7):1644–1668, 2008.

Amy Finkelstein, Nathaniel Hendren, and Erzo F.P. Luttmer. The value of medicaid: Interpreting results from the oregon health insurance experiment, 2014.

Mikhail Golosov and Aleh Tsyvinski. Designing optimal disability insurance: A case for asset testing. *Journal of Political Economy*, 114(2):257–279, 2006.

Benjamin R Handel. Adverse selection and switching costs in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–2682, 2013.

Benjamin R Handel, Igal Hendel, and Michael D Whinston. Equilibria in health exchanges: Adverse selection vs. reclassification risk. 2013.

R. Glenn Hubbard, Jonathan Skinner, and Stephen Zeldes. Precautionary saving and social insurance. *Journal of Political Economy*, 103(2):360–399, 1995.

Kaiser Family Foundation. State health facts, total hospitals, April 2011. URL `http://kff.org/other/state-indicator/total-hospitals/`.

Ahmed Khwaja. Estimating willingness to pay for medicare using a dynamic life-cycle model of demand for health insurance. *Journal of Econometrics*, 156(1):130 – 147, 2010.

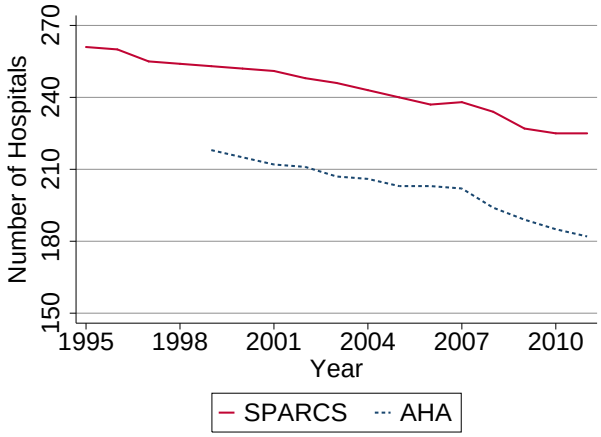
Amanda Kowalski. The early impact of the affordable care act state-by-state. *Brookings Papers on Economic Activity*, 2014.

Mark V. Pauly, Howard Kunreuther, and Richard Hirth. Guaranteed renewability in insurance. *Journal of Risk and Uncertainty*, 10(2):143–156, 1995. ISSN 0895-5646. doi: 10.1007/BF01083557. URL `http://dx.doi.org/10.1007/BF01083557`.

A Data

Our main data are restricted use data from the Statewide Planning and Research Cooperative System (SPARCS), which contain data on almost all inpatient discharges in New York state from 1995 to 2011 — a period of 17 years. SPARCS data include all inpatient discharges from all Article 28 licensed hospitals, defined as general hospitals, nursing homes, and diagnostic treatment centers. To examine the completeness of the SPARCS data, we compare it to New York state hospital counts from the American Hospital Association (AHA) Annual Survey, as compiled by the Kaiser Family Foundation (Kaiser Family Foundation, 2011). We plot the number of hospitals from both sources in Figure A1. There are approximately 225 to 250 hospitals in the SPARCS data in each year from 1995 to 2011. The figure shows that our data include *more* hospitals in each year than AHA reports because the AHA Annual Survey only includes "community hospitals," a smaller subset of all hospitals.

Figure A1: Hospital Count Comparison to the AHA



Combining the SPARCS data with other sources, we construct a balanced panel designed to represent all individuals in the state of New York from 1995 to 2011 who were ages 25-64 in 1995. We focus on individuals ages 25-34 in 1995 to isolate individuals who are young but working-age. We further restrict our sample to allow us to focus on young individuals for whom public health insurance has value *only* as a potential safety net. Our main sample, the "private in 1995 sample," includes all individuals ages 25-34 in 1995 with private coverage in 1995. This balanced sample includes annual insurance coverage and spending for 1,689,078 individuals for 17 years, 1,199,245 of whom appear in a hospital at some point in our data. We give an overview of how we construct our data here, and we provide full detail in the data appendix.[9].

To create our balanced panel, we first create an observation in every year of our sample period for every individual ages 25–64 in 1995 who appears in our SPARCS data in any year. Since we only observe individuals in the years in which they show up in a hospital, this step ensures that we

[9]The data appendix for this paper is available at `http://www.econ.yale.edu/~ak669/longitudinal.data.appendix.pdf`

have a balanced panel for the individuals who ever appear. However, many individuals never show up in a hospital during our sample period. Therefore, we create observations for the remaining individuals who were ages 25–64 in 1995 and living in New York state, according to New York state population totals and death rates taken from the Current Population Survey (CPS). We obtained special restricted-access linkage of our SPARCS data to state Vital Statistics data; we therefore observe deaths, even if they occur outside of the hospital.

Next, we assign inpatient costs and insurance coverage within our balanced panel. Assuming that we observe all inpatient costs in the state of New York, costs are equal to what we observe in the data and equal to zero otherwise. We observe hospital *charges* (list prices) rather than costs in SPARCS. However, we follow standard practice and calculate costs by deflating observed charges by the cost-to-charge ratio reported to Centers for Medicare & Medicaid Services (CMS) for each hospital-year. We implement this process at the individual-discharge-year level prior to aggregating the data to the individual-year level to account for variation in the cost-to-charge ratio across hospitals. We further deflate hospital costs by the CPI-U so that we always express costs in 2012 dollars.

Finally, we define an upper bound for annual individual costs of $M = \$30,000$ (in \$2012) to account for the ability of insurers and individuals to negotiate discounts with hospitals in the event of exorbitant medical costs. We use the same upper bound for all individuals in all periods, regardless of insurance status. Approximately 0.4% percent of individuals in our private in 1995 sample exceed this upper bound in each year. We examine robustness to this assumption and find that the main results hold for a wide range of values of M.

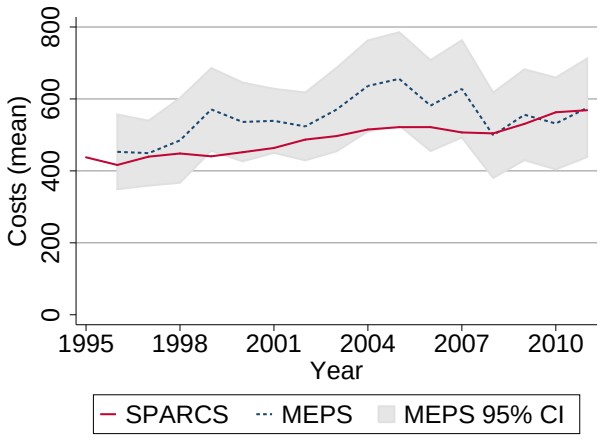
The main selection issue in our data is that we do not observe if and when individuals exit the state of New York during our sample period. As long as individuals do not differentially move out of the state of New York after a health event, attrition is much less likely to be a problem in our data than it is in other data sets. Furthermore, attrition will bias us *against* finding a larger longitudinal value of insurance relative to the value of insurance that we find in repeated cross-sections.

An additional caveat is that results based on data from the state of New York might not generalize to other states. During the entire period of our sample, New York had strict regulations in its individual health insurance market (the market for health insurance purchased directly through insurers). Since before the start of our sample in 1993, it has had "community rating" regulations that limit premium variation across beneficiaries and "guaranteed issue" regulations that require insurer to offer insurance to all applicants. These regulations have been established nationally by the ACA in 2014, such that results from New York might be of general interest. However, our analysis covers all forms of health insurance, of which individual health insurance is only a small part, currently covering less than 2% of individual in the state of New York (Kowalski, 2014). Results from New York might not generalize for other reasons, which would be hard to separate empirically. However, we compare our data to data from nationally-representative surveys to gauge its external validity.

Inpatient Cost Comparison to the MEPS We compare inpatient costs in our data to inpatient costs in the MEPS. Within the private in 1995 sample, average annual costs after imposing

M are \$489, ranging from a low of \$416 in 1996 to a high of \$569 in 2011. As exhibited in Figure A2, the pattern of costs over time follows a very similar trajectory in both MEPS and SPARCS from 1996 (the first year of the MEPS) onward. In MEPS, average annual inpatient costs are \$549, ranging from a low of \$449 in 1997 to a high of \$656 in 2005. However, due to the substantially smaller sample size, the MEPS statistics are less stable over time than the SPARCS statistics, as shown by the 95% confidence intervals calculated within the MEPS. For example, in the 2011 MEPS survey, there are only 4,499 individuals who were between the ages of 25 and 34 in 1995, which means the MEPS is approximately 0.3% as large as our SPARCS sample.[10] We do not report confidence intervals for the SPARCS data because they are not weighted — they include the universe of individuals that meet the sample inclusion criteria. Among only those members of the private in 1995 sample that visit the hospital, average annual costs after imposing M is \$8,028. The comparable figure in the MEPS is \$10,029; however, again, due to the small sample size in the MEPS, this estimate is not as precise as the SPARCS estimate.

Figure A2: Inpatient Cost Comparison to the MEPS



A major concern with our SPARCS data is that we only observe inpatient costs. From the MEPS data, we see that inpatient costs constitute a fairly large share of total costs over the period. The ratio of inpatient costs to the sum of inpatient and outpatient costs, ranges from a high of 73% in 2000 to a low of 58% in 2009, generally declining throughout the period. We restrict our analysis to inpatient costs because we are not aware of any other data, including the MEPS, which would allow us to observe longitudinal patterns in outpatient costs for a similar population for 17 years.

Insurance Coverage Comparison to the CPS We divide health insurance coverage into three categories: private, public, and uninsured. Public health insurance includes Medicaid coverage, Medicare coverage, and other types of public coverage. It is straightforward to assign insurance

[10]To be as consistent as possible with our SPARCS private-in-1995 sample, we restrict our MEPS sample to individuals who were ages 25-34 in 1995 and privately insured in the current year. In addition, we deflate cost data in the MEPS by the CPI-U so that all costs are in 2012 dollars and apply M as an upper bound to individual costs in any given period.

coverage for people that we observe in the hospital — we simply assign the observed coverage for the given year (if we observe multiple coverage types within the same year, we assign coverage to be the first observed type of coverage for that year). For individuals that we do not observe in the hospital in 1995, we randomly assign insurance coverage so that our distribution of insurance types reflects the CPS coverage estimates for our age cohorts of interest in New York in 1995. In remaining periods for which we do not observe coverage, we assign coverage to be the same type as assigned previously. Although this methodology under-counts changes in coverage in our data, the only coverage changes that we do not observe are for individuals who have zero inpatient hospital costs in the corresponding year. Therefore, we capture the most important changes in health insurance coverage because we observe health insurance coverage for all individuals who visit the hospital.

Figure A3 shows how coverage ranges from the CPS[11] (top left) and from the SPARCS data alone (top right) combine to achieve coverage in our private-in-1995 sample (bottom). In the CPS, most individuals have private insurance, and the rate of private insurance remains fairly stable across the period but varies somewhat from year-to-year, from a minimum of 61% in 1996 to a maximum of 71% in 2007. The rate of public insurance also remains quite stable, but increases gradually over the period. The lowest rate of public insurance is 9.7% in 2000, and the highest is 17.8% in 2011. The rate of uninsurance declines somewhat throughout the period. The highest rate of uninsurance is 25.8% in 1996, and the lowest is 13.6% in 2007.

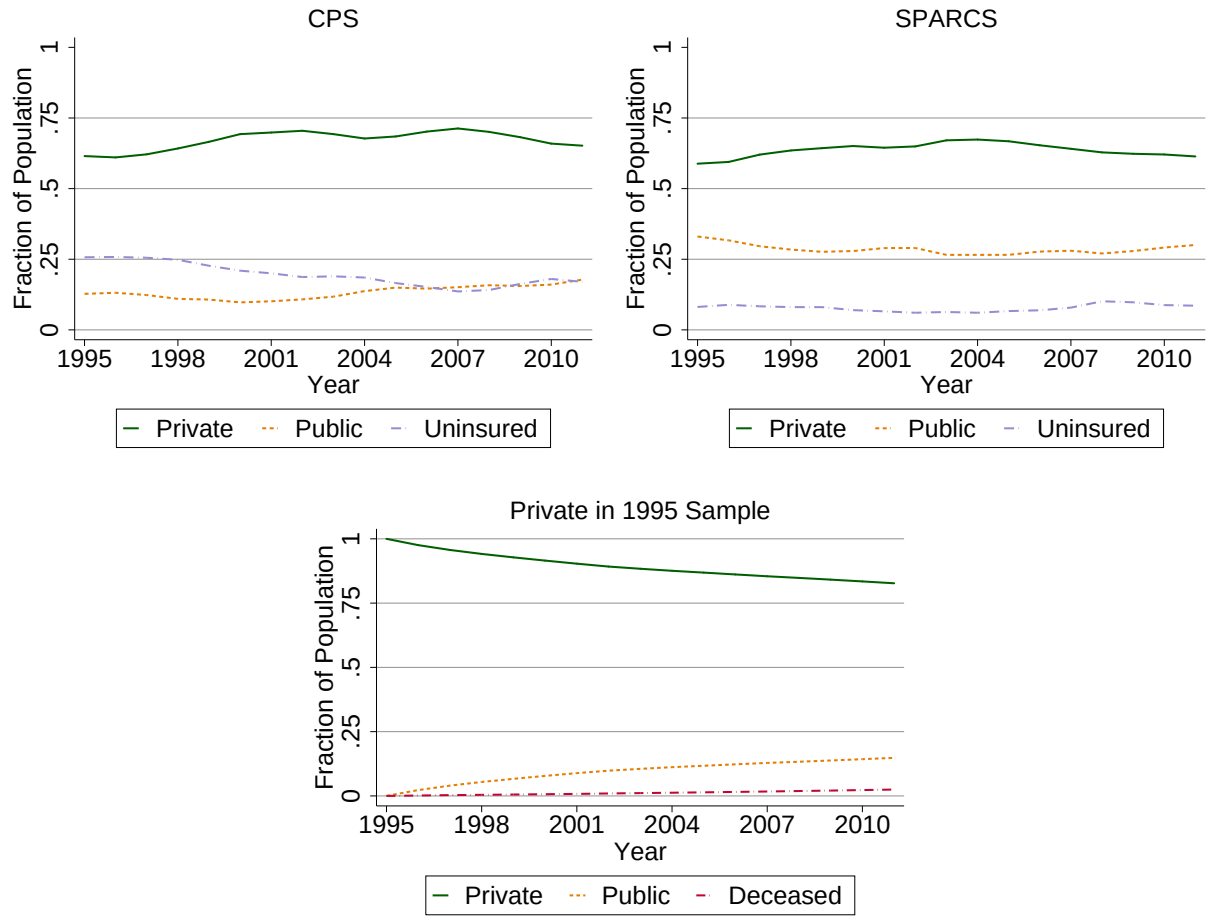
We see similar insurance rates in SPARCS, though we observe somewhat higher rates of public insurance and lower rates of private insurance and uninsurance. These differences are not surprising, given that we are observing only inpatient hospital visitors, and inpatient hospital visitors will be more likely to have coverage if there is adverse selection into insurance. The fact that public insurance rates are higher within the SPARCS inpatient population than within the population as a whole motivates our interest in public insurance as a safety net for the very sick.

The bottom graph shows coverage in the private-in-1995 sample. Per our focus, 100% of this sample has private coverage in 1995, but some members of the sample gain public coverage over time and others die. In our private-in-1995 sample, we exclude individuals who ever become uninsured because doing so simplifies the intuition behind our calculations. After we build up the intuition and the main results, we add individuals who ever become uninsured to the private-in-1995 sample to construct our preferred sample in Section 5. As shown, the proportion of the private-in-1995 sample that has public coverage grows throughout the period, motivating the role of public health insurance as a safety net for this population. As shown in Figure A3, by 2011, 2.5% are deceased and 14.8% are publicly insured. Although the rate of public health insurance in the private-in-1995 sample is not as large as the rate of public health insurance in the full CPS population, which includes those who started out with public health insurance or uninsured, we can see that a significant portion of individuals with public health insurance come from the ranks of the previously privately insured.

We utilize the longitudinal nature of our private-in-1995 sample in determining the longitudinal value of public insurance. For comparative purposes, we can obscure the longitudinal nature of our private-in-1995 sample by simply omitting the longitudinal identifiers. Such cross-sectional

[11] CPS estimates are simple (unweighted) two-year moving averages of the previous year with the current year.

Figure A3: Insurance Coverage Comparison to the CPS



data constructed from our data can then offer a cleaner comparison between our data and cross-sectional survey data such as the CPS, because it holds features like sample size constant. Cross-sectional data derived from our longitudinal data are arguably superior to traditional cross-sectional surveys in the sense that our data "samples" the same individuals each year, thereby controlling for demographic variables in each year. Our cross-sectional data also only "samples" individuals who had private insurance in 1995, thereby providing more information than a true cross-sectional survey. Nevertheless, we show that it still offers a very limited picture of the value of public health insurance as a safety net relative to our full longitudinal data.