The Relationship Between Income, Health Status, and Health Expenditures in the United States

By

David Bueno

Ph.D. Electrical and Computer Engineering
University of Florida, 2006

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FUFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF BUSINESS ADMINISTRATION
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2011

Signature of Author: _____
MIT Sloan School of Management
May 6, 2011

Certified By: _____
Vivek Farias
Assistant Professor, MIT Sloan School of Management
Thesis Supervisor

Accepted By: _____
Debbie Berechman
Executive Director, MBA Program
MIT Sloan School of Management

The Relationship Between Income, Health Status, and Health Expenditures in the United States

By

David Bueno

Submitted to the MIT Sloan School of Management on May 6, 2011 in partial fulfillment of the requirements for the degree of Master of Business Administration

ABSTRACT

The relationship between income and health has important implications for policy makers and businesses, and will continue to receive attention as healthcare reform takes hold in the U.S. Most existing literature looks at the relationship between income and either health status or health expenditures in isolation. However, in this research, we take advantage of the wealth of data available in the U.S. Department of Health and Human Services' Medical Expenditures Panel Survey (MEPS) to answer two important, related questions regarding the income-health relationship for U.S. adults. First, we seek to determine how much sicker are poorer people than richer people (if at all), both in their perception and in actual terms. Second, we seek to determine if a poorer person is likely to consume more or less care than a richer person for given level of health or condition.

To answer the first question, we start by examining the relationship between family income and health status using multiple regression techniques. For both perceived health and actual health, we find a curvilinear relationship between income and health, with diminishing returns associated with membership in successively higher-income groups. Depending on the status metric, the associated health benefits of membership in high-income cohorts tend to flatten once income reaches approximately 500-600% of the federal poverty level (FPL). We also find that marginal income at low income levels tends to be more strongly associated with reduced probability of poor health than increased probability of strong health. Regardless of the dependent variable chosen, we find that the shape of the relationship between income and health status is the same once we normalize the coefficients. Perceived and actual health are strongly related, although some of our results indicate that poorer people may be more pessimistic about their health than richer people.

We find similar trends when we examine the relationship between income and health expenditures using the MEPS data. In this case, however, the diminishing returns associated with membership in higher-income cohorts are more accelerated, and the associated reductions in spending for membership in successive cohorts above 200-300% FPL are not significantly different from zero. When we add controls for health status, however, we find that the wealthiest members of the population are most likely to have the highest spending on healthcare, although not drastically so. In addition, we find the poorest members of the population do not have a tendency to overconsume care relative to their level of health.

Thesis Supervisor: Vivek Farias
Title: Assistant Professor, MIT Sloan School of Management

Table of Contents

Table of Tables

Table of Figures

1. Introduction

Despite decades of research into the relationship between income and health in the U.S. and across the world, the exact nature of the relationship is still not completely understood. Much of the interest in the income-health relationship traditionally results from the fact that it plays an important role in policy decisions of governments. If low income is found to cause poor health, and if the marginal returns to income at high income levels are diminishing, some form of income redistribution could in fact improve the overall health of a population. In addition to aiding governments, a better understanding of the income-health relationship could also help employers better predict the health expenditures of segments of their employee bases, or could help insurance payers better anticipate expenditures for new or existing client segments. Especially in light of the uncertainty revolving around U.S. healthcare reform, any additional information on the expected health of potential client segments could be extremely useful to U.S. payers.

The income-health relationship is quite complex, however, and previous studies have found varying and sometimes conflicting results. These studies are complicated by the subjectivity of health measurements and the variety of possible income metrics to choose from. It is also quite difficult to determine the magnitude and direction of causality, even when a relationship can be proved, since income affects health, but health also affects income (e.g., since sicker people may be less able to work).

Much of the prior work in the area of the income-health relationship involves absolute income effects and income inequality effects. Absolute income effects measure the relationship between income and the health of an individual or group (family, population, etc.). Income inequality effects, on the other hand, focus on the relationship *between* incomes within a population. Populations with high income inequality are ones where the top earners earn a disproportionate share of the population's total income. Income inequality has been shown by many sources to be increasing in the U.S. since 1968, reversing an earlier decreasing trend (U.S. Census Bureau 2000) and the U.S. has one of the largest levels of income inequality among high-income countries as measured by the commonly referenced Gini index (CIA 2010).

It is widely accepted that there is a direct relationship between absolute income and health, but the exact shape and nature of the relationship is still not clear. This relationship is difficult to determine, with many possible measures of both income and health available. Most researchers, especially those studying the U.S., believe the relationship is curvilinear, with diminishing returns to high income (McDonough 1997; House 1990). However, other studies, such as one in Finland, have reported nearly linear associations between income and health (Der 1999). However, assuming a curvilinear relationship exists, individual-level data is best to study (rather than averaged population data), since measures of average income or health

in a community can mask the true relationship between income and health. This makes comparisons across societies very difficult, since most large individual-level surveys that gather health and income data focus on a particular country.

Many researchers believe that income inequality also specifically impacts health. This relationship is usually cited in addition to the absolute income effects. However, this relationship is even more difficult to ascertain than the absolute relationship. One challenge revolves around the curvilinear findings described above. If the poorest segments of the population are most likely to have poor health, it follows that populations with high levels of income inequality will also have worse health than those with a more even income distribution, all other factors being equal. This trend occurs because the lowest-income groups in the population contribute disproportionately to the overall decrease in the health of a society, while the highest-income groups contribute less significantly to the overall increase in the health of a society due to diminishing returns to health at high levels of income. However, some research indeed shows distinct impacts on health resulting from income inequality that are separate from the curvilinear absolute income effects (Wagstaff 2000). Many studies use averaged data to show a relationship between income inequality and poor health, however, and these studies are incapable of distinguishing between income inequality effects and curvilinear absolute effects.

The scope of this paper will focus on the absolute relationship between income and health in the U.S. for adults age 18 and up. The absolute income-health relationship is the more accepted and understood of the two described above, and arguably a much larger driver of health status for U.S. individuals than income inequality. However, this work will take a unique approach to this topic, as it will seek to explore not only the relationship between income and health *status*, but also explore the relationship for the same population between income and health consumption. First, we seek to determine whether poorer people are likely to be healthier or sicker than richer people, but in perception and reality. Then, we will also determine whether a poorer person is likely to consume more or less care than a richer person for the same condition or relative level of health. While we will touch on causality where appropriate, we are more concerned with the magnitude and shape of the association between income and health. Therefore, the research will be more useful towards users that want to learn more about the expected health characteristics of a particular segment than users that want to make policy decisions regarding income redistribution and public wellbeing.

The remainder of this paper is organized as follows. Section 2 provides an overview of the Medical Expenditures Panel Survey (MEPS) whose data our study is based upon. It also describes the key dependent variables that we derive from the MEPS data for use in our models. Section 3 then explores the relationship between income and health status to determine how much sicker poorer people are likely to be than richer people, in perception and reality. In Section 4, we explore the relationship between income and healthcare consumption. This section first examines the relationship between income and healthcare consumption using an approach

similar to the one used for health status. Then, we add a control for health status, in order to explain the relationship between income and health expenditures, *for individuals of similar actual health.* In other words, we seek to explain whether low-income individuals are more or less likely to consume large amounts of health care than high-income individuals for a similar condition. Section 5 presents a summary of this work and suggestions for future research, followed by references.

2. Medical Expenditure Panel Survey and Data Overview

The Medical Expenditure Panel Survey (MEPS) consists of several surveys of families and individuals, medical providers, and employers in the U.S. MEPS is conducted by the Agency for Healthcare Research and Quality (AHRQ), which is part of the U.S. Department of Health and Human Services (HHS). While there are several MEPS surveys conducted each year, our research will exclusively use data from the MEPS Household Component survey, which "fields questionnaires to individual household members and their medical providers to collect nationally representative data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment." (U.S. Dept. of HHS 2010) In this section, we provide an overview of the MEPS Household Component, discuss the data that we derived from the MEPS survey, and briefly discuss some MEPS data that we have excluded from our analysis.

2.1. MEPS Household Component Overview

The MEPS Household Component (MEPS-HC) is a nationally representative sample of the U.S. civilian, noninstitutionalized population. The survey is administered in-person by a trained interviewer via computer-assisted personal interviewing (CAPI), which uses a laptop computer to aid in data collection. Each 2007 MEPS participant is interviewed three times in 2007, and many questions (e.g., most health status questions) are asked in each round, with responses recorded separately. We use the 2007 MEPS HC-113 dataset, which is the most recent survey publicly available as of September 2010 that contains data regarding individual incomes and medical expenditures (released in November 2009). The sample is drawn from the National Health Interview Survey, and the set of valid MEPS-HC samples for 2007 ultimately consists of 29,370 persons and 11,615 families. Data is indexed on a person-level basis, and person-level weights are included with the dataset such that estimates can be made for the entire civilian noninstitutionalized U.S. population for 2007.

2.2. Data Overview

For this research, we are most interested in the data on demographics (particularly income, but also other relevant controls), health status, healthcare consumption, and insurance status.

Table 1 contains a description of the indicators used in later sections and explains how they were derived from the MEPS-provided variables, when applicable.

Table 1- Description of Indicators Derived from MEPS Variables

Indicator category	Indicator type	Indicator description	Indicator categories
Demo-graphics	Income	Indicator for a range of family income as a percentage of the 2007 U.S. federal poverty level (FPL) for a family of that size. Each range is inclusive of the lower bound and exclusive of the upper bound.	<100% FPL 100-133% FPL 133-200% FPL 200-300% FPL 300-400% FPL 400-500% FPL 500-600% FPL 600+% FPL
	Age	Indicator for the age range of the individual on December 31, 2007.	18-25 years 26-35 years 36-45 years 46-55 years 56-65 years 66-75 years 76+ years
	Race/eth-nicity	Indicator for the race (or ethnicity) of the individual.	Black Asian Hispanic Other
	Gender	Indicator for the gender of the individual	Male/female
Health status	Perceived health	Indicator for the perceived health of the individual in each of the three MEPS-HC rounds. The respondent gives a unique assessment in each round, and there are five health categories, so there are 15 total indicators.	Excellent Very good Good Fair Poor
	Strong health	Indicator that the individual was perceived to be in very good or excellent health in each of the three MEPS rounds.	Yes/no

	Weak health	Indicator that the individual was perceived to be in fair or poor health in each of the three MEPS rounds.	Yes/no
	Perceived health index	Sum of the individual's perceived health scores for the three MEPS rounds.	N/A (integers 0-16)
	MEPS "priority conditions"	Indicator for each of 11 MEPS "priority conditions," reflecting that a physician has diagnosed the individual with the condition at some point in the individual's life.	High blood pressure Coronary heart disease Angina Heart attack or myocardial infarction Other heart disease Stroke Emphysema High cholesterol Diabetes Arthritis Asthma
	Count of priority conditions	Number of MEPS priority conditions that the individual has ever been diagnosed with. For count purposes, treats all four forms of heart disease as a single condition (maximum eight total conditions).	N/A (count range 0-8)
	Instru-mental daily living activity limitations	Indicator stating whether the individual received help with "instrumental" daily activities such as using the telephone, paying bills, taking medications, preparing light meals, doing laundry, or going shopping.	Yes/No
	Other daily living activity limitations	Indicator stating whether the individual received help with daily activities such as bathing, dressing, and getting around the house.	Yes/No

	Functional/ activity limitations	Indicator stating whether the individual had difficulty walking, climbing stairs, grasping objects, reaching overhead, lifting, bending or stooping, or standing for long periods of time.	Yes/No
	Assistive technology use	Indicator stating whether the individual used aids such as a walker, bars in the bathtub, or other special equipment for personal care or everyday activities.	Yes/No
Healthcare consump-tion	Health expendit-ures	Total health expenditures in 2007 as defined by MEPS.	N/A (continuous)
Insurance status	Insurance status	Indicator stating whether the individual had any health insurance coverage at any point in 2007.	Yes/No

While most of the indicators in

Table 1 are self-explanatory and will also be discussed in more detail later in this document, several warrant further description here.

Due to the expected curvilinear relationships between income and age (as independent variables) and health (as a dependent variable), binary indicators for income and age are commonly used in existing literature (McDonough 1997; House 1990). We use the 2007 U.S. federal poverty guidelines rather than raw income in order to efficiently account for family size in our income measures. Our income range indicators were chosen based on preliminary analysis of the MEPS HC-113 data with a goal of covering the full spectrum of income levels with a reasonable number of indicators. We separate the 100-133% FPL range for one important reason: the U.S. healthcare reform Patient Protection and Affordable Care Act (PPACA) expands Medicaid eligibility in all states to cover individuals with income up to 133% FPL (The Kaiser Family Foundation, 2010). In addition, this group may migrate in and out of the subsidized health insurance exchanges as their income shifts up and down, so it is important to both insurers and public policymakers to understand the health characteristics of this population sub segment. We observed almost no statistical difference in the health or spending activity between individuals above ~500-600% FPL in our preliminary analyses, making 600% FPL our selected upper income threshold. MEPS caps reported age at 85 years in order to preserve anonymity in their data, so 76+ years is a natural top-end age range for our study. We also include indicators for race/ethnicity and gender. We include white individuals in the "Other" category.

Because health status is very subjective and difficult to measure, we include several measurements of health status in our various empirical analyses. One measurement provided by MEPS is perceived health. MEPS asked the survey taker to provide perceived health status for each family member as either poor, fair, good, very good, or excellent three times during the 2007 survey year. To maintain maximum information for relevant analyses in which health status is an *independent* variable, we use binary indicators for each of the three rounds. For this purpose, we include the very small number of invalid ("don't know", "refused", or "not ascertained") responses in the "poor" health category. However, we also derive an indicator for "strong health" explaining whether an individual was perceived to be in either "very good" or "excellent" health for each of the three survey rounds. The benefit of this approach is that it allows for some natural variance in an individual's responses from round to round (despite that their health may be relatively constant), and adjusts somewhat for the fact that different individuals will have a different idea of what "very good" or "excellent" health is. However, we expect that most generally healthy individuals would select one of the top two out of the five possible responses. For similar reasons, we derive an indicator for "weak health," which is positive when an individual is perceived to be in "fair" or "poor" health in each of the three rounds (for this indicator, we exclude individuals with any invalid responses). Finally, to incorporate maximum information into a single indicator, we calculate a "perceived health index" with the sum of the individual's scores for the three rounds. This index assigns 0-4 points for each assessment from poor to excellent, respectively, such that the maximum score is 16 and the minimum score is 0 (invalid responses also yield 0 points). We will discuss the benefits and limitations of the "health index" approach later in this paper.

The MEPS-HC survey also asked individuals whether they had been diagnosed with each of several common "priority conditions" listed in

Table 1. MEPS also includes "joint pain" as a priority condition, however, we omit this condition because we feel it is more inherently more subjective than the other conditions. The other conditions each require a doctor's diagnosis to answer affirmatively, while joint pain is a pure self-assessment in the MEPS survey. A positive value for our indicator states that the individual has been diagnosed with the condition (and hence, negative responses and invalid responses are treated the same way). We also calculate a count of the number of priority conditions with which an individual has been diagnosed. For this count, we treat all forms of heart disease as a single diagnosis to avoid double counting of related heart ailments.

In addition to priority conditions, MEPS asks about "instrumental" daily living limitations, other daily living limitations, activity limitations, and use of assistive technology. These four indicators are also described in

Table 1. We set these indicators to a value of 1 if the individual is described as having the limitation during any of the rounds in which the question was asked. Otherwise, the indicator value is set to 0.

MEPS also provides a variable describing total expenditures for healthcare services for an individual over the course of the year. This figure accounts for expenses paid to healthcare providers by the individual (note that this does not include insurance premiums), his/her insurer, and the government.[1] Our indicator uses the dollar value provided in this MEPS variable directly. In order to increase the accuracy of expenditure data, the MEPS Household Component (HC) survey uses data from the MEPS Medical Provider Component (MPC) survey to augment the expenditure data (U.S. Dept. of HHS 2009).

We also include an indicator stating whether the individual had insurance at any point in 2007 (no invalid responses for the associated variable are present in the HC-113 dataset).

We have intentionally excluded geography and education level as independent variables in our analyses. We do not include geography because MEPS does not provide detailed geographic information to the public due to confidentiality reasons. MEPS only provides four census-level regions (Northeast, Midwest, West, South), which do not provide enough granularity to use in our models. We exclude education level due to high collinearity with income. For public policy or business decision-making purposes, separation of the income and education effects on health status would provide limited, if any, additional benefit beyond simply considering income.

3. Income and Health Status

In this section, we seek to determine how much sicker (if at all) poorer people are likely to be than richer people. To that end, we first investigate the relationship between income and various measures of perceived health. We then seek to understand the relationship between income actual health. We conclude the section by contrasting the relationships between income and the various health measures and then analyzing the relationship between perceived health and actual health across income levels.

---

[1] As described in the MEPS HC-113: 2007 Full Year Consolidated Data File: "expenditures in MEPS are defined as the sum of direct payments for care provided during the year, including out-of-pocket payments and payments by private insurance, Medicaid, Medicare, and other sources. Payments for over-the-counter drugs are not included in MEPS total expenditures. Indirect payments not related to specific medical events, such as Medicaid Disproportionate Share and Medicare Direct Medical Education subsidies, are also not included." (U.S. Dept. of HHS 2009)

3.1. Income and Perceived Health

During each of the three MEPS-HC rounds, the survey asks the respondent to assess the health of each individual in the family as excellent, very good, good, fair, or poor, compared to other individuals of the person's age. While perceived health is clearly subjective, we believe it provides a directionally strong measure of an individual's overall health and is an appropriate starting point for a discussion on income and health status. In order to help mitigate the subjectivity of perceived health status, we use the "strong health" and "weak health" indicators that we derived from the MEPS data using the methodology explained in Section 2.2. These indicators describe consistently strong ("very good" or "excellent") or consistently weak ("fair" or "poor") health assessments for an individual in each of the three MEPS-HC rounds. For these analyses, we are interested in determining the probability of reporting strong or weak health, and hence, a logistic regression is an appropriate model. We run separate regressions for strong health and weak health, with the dependent variable as a binary indicator for strong or weak health. A vector of income variables is used to describe family income range, with 600+% FPL as the omitted category to test the relative differences in expected perceived health for other income groups versus this highest-income group. For some regressions, we also include a set of control variables. All dependent and independent variables used in this paper were described in

Table 1, although this analysis (and, in fact, all of our other analyses) uses a subset of the variable list. In this case, our full set of controls includes age, race/ethnicity, and gender. The age segment from 18-25 is the omitted age category, and "Other" is the omitted race/ethnicity category (which includes white races). We employ weighted regression procedures using Stata/SE 11.1 throughout this work to make use of the MEPS HC-113 sampling weights for the noninstitutionalized civilian U.S. population. This approach allows us to use the MEPS data to report accurate results based on the entire relevant population.

Table 2 shows the raw results of the logistic regression model relating perceived strong health and income. The regression coefficients are the natural logarithm of the odds ratio of reporting strong health relative to the omitted category. For ease of interpretation, we then numerically translate the logistic coefficients into the *probabilities* of each income cohort reporting strong health, with all other factors held constant at their baseline level (i.e., the omitted category for each control), and these results are shown in Table 3. Table 4 and Table 5 show the analogous results for the weak health/income model. In each table we show results for models with no controls, age controls only, and the full set of age, gender, and race/ethnicity controls. For conciseness, only the income coefficients and associated statistics (rather than all coefficients) are included in the tables. It is also important to note that weighted regression procedures such as those used in this paper do not typically produce correct classical standard errors (Gelman 2007). Therefore,

throughout this paper, we report the robust standard error output from Stata unless otherwise noted.

Table 2- Relationship Between Income and Perceived Strong Health

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	-1.2130 (.0648)**	-1.4428 (.0676)**	-1.3319 (.0695)**
100-133	-1.2447 (.0852)**	-1.3576 (.0881)**	-1.2613 (.0895)**
133-200	-.9037 (.0626)**	-1.0409 (.0656)**	-.9582 (.0668)**
200-300	-.5615 (.0558)**	-.7365 (.0585)**	-.6778 (.0593)**
300-400	-.4818 (.0595)**	-.5969 (.0616)**	-.5674 (.0620)**
400-500	-.2749 (.0648)**	-.3941 (.0664)**	-.3776 (.0669)**
500-600	-.0677 (.0702)	-.1392 (.0719)	-.1282 (.0723)
Constant	.1232 (.0379)**	.9639 (.0639)**	1.0243 (.0685)**
Pseudo R^2	.0299	.0603	.0650

*Significant at 5% level; **Significant at 1% level.

Table 3- Probabilities of Reporting Perceived Strong Health

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	0.2516	0.3825	0.4237
100-133	0.2457	0.4028	0.4410
133-200	0.3142	0.4808	0.5165
200-300	0.3921	0.5566	0.5858
300-400	0.4113	0.5907	0.6123
400-500	0.4622	0.6387	0.6562
500-600	0.5139	0.6953	0.7101
600+	0.5308	0.7239	0.7358
Mean of dependent variable	0.4117	0.4117	0.4117

As expected, Table 2 and Table 3 show a positive correlation between strong perceived health and high income. As mentioned previously, 600+% FPL is the omitted category, and coefficients presented in Table 2 indicate the log of the odds of each income category reporting strong health divided by the odds of the 600+% FPL category reporting strong health. However, the absolute probabilities shown in Table 3 are much more intuitive. Regardless of controls, the relationship appears to be strongly curvilinear, with diminishing positive returns associated with membership in higher-income groups once income is sufficiently above the federal poverty level. In all cases, there is no statistically significant change in perceived strong health between the 500-600% FPL group and the 600+% FPL reference group. In the "No Controls" case, there is actually a small and statistically insignificant decrease in perceived health associated with income in the 100-133% FPL range compared to income below the poverty line. Once controls are added, there are small improvements to health associated with membership in the 100-133% FPL group relative to the <100% FPL group. However, these relatively small differences suggest that there may be a minimum income threshold near the poverty line at which the association between income and perceived strong health is at its strongest (at which point, we then see a weakening of the perceived health benefits associated with income as income increases). The coefficients of the controls are as expected, with race/ethnicity and gender indicators possessing relatively low-magnitude coefficients (no larger than -.3908), and age indicators possessing relatively high-magnitude coefficients (up to -1.5661 for the oldest age group with full controls) despite that the MEPS-HC survey asks respondents to state perceived health in comparison to others of the same age.

Table 4- Relationship Between Income and Perceived Weak Health

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	1.9510 (.1358)**	2.3234 (.1387)**	2.2972 (.1426)**
100-133	1.7440 (.1565)**	1.9561 (.1634)**	1.9176 (.1664)**
133-200	1.3095 (.1460)**	1.5425 (.1486)**	1.5172 (.1509)**
200-300	.9931 (.1428)**	1.2925 (.1443)**	1.2711 (.1456)**
300-400	.8314 (.1598)**	1.0166 (.1606)**	.9984 (.1607)**
400-500	.3530 (.1841)**	.5397 (.1848)**	.5224 (.1848)**
500-600	.2097 (.2172)**	.3417 (.2178)**	.3329 (.2182)**
Constant	-3.7949 (.1213)**	-6.0457 (.2365)**	-5.9782 (.2423)**
Pseudo R^2	.0516	.1214	.1236

*Significant at 5% level; **Significant at 1% level.

Table 5- Probability of Reporting Perceived Weak Health

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	0.1366	0.0236	0.0246
100-133	0.1140	0.0165	0.0169
133-200	0.0769	0.0110	0.0114
200-300	0.0572	0.0086	0.0089
300-400	0.0491	0.0065	0.0068
400-500	0.0310	0.0040	0.0043
500-600	0.0270	0.0033	0.0035
600+	0.0220	0.0024	0.0025
Mean of dependent variable	0.0565	0.0565	0.0565

Similarly to our results for strong health, the results in Table 4 and Table 5 show diminishing returns with respect to perceived weak health (or rather, the lack of perceived weak health) associated with membership in higher-income cohorts. The magnitudes of the coefficients are much smaller than those for strong health, due to the much smaller percentage of survey respondents reporting fair or poor health in all three rounds. However, these results actually indicate that poverty has a larger association with the chance of reporting weak health than it does with the chance of reporting strong health. For example, in the case without controls, an individual with income below the poverty line is 6.2x more likely to report weak health than an individual with family income 600% FPL or above. This observation is in contrast to the strong health scenario, where an individual with family income 600% FPL or above is only 2.1x more likely to report strong health than someone whose family income is below the poverty line. Also in contrast to the results for strong health, these results show very strong improvements associated with income in the 100-133% FPL range versus the range below the poverty line. This finding possibly indicates (assuming a causal relationship) that the marginal returns of income in terms of *preventing very bad health* in low income ranges are stronger than income's marginal returns in terms of *facilitating good health*. Age again has the strongest influence of controls, with race/ethnicity and gender exhibiting a much weaker effect.

While our results for weak and strong health are useful because they are easily understood and allow for reasonable variance in individual reporting, they are also limited because they discard information that we can incorporate from the full dataset. The discarded information is both temporal and spatial, because our dependent variables only cover individuals who respond similarly for each of three rounds. To help incorporate this missing information into our analyses, we also derived a "perceived health index" as explained in Section 2.2. This index ranks the perceived health of the individual across all three MEPS-HC rounds, with a maximum score of 16 (for someone who responded "excellent" in each round) and a minimum score of 0 (for someone who responded "poor"" in each round). We then perform an OLS linear regression with parameters similar to those used in the logistic regression previously, with the exception that the dependent variable is now

the perceived health index. Therefore, the model is no longer a logistic model with binary output, but rather one that captures the associated change in the index as income varies from the omitted 600+% FPL category, with an index change of magnitude 1.0 indicating an improvement (or decline) of one response level (e.g., "excellent" vs. "very good") in one round of the survey.

While this approach incorporates data obtained in the MEPS-HC survey that we could not integrate into the logistic models for strong or weak health, its drawback is that it inherently assumes each perceived health-ranking level is of equal value in developing the index. For example, the model assumes that the increment from "poor" to "fair" health is no more or no less important than the increment from "very good" to "excellent" health. Ordered probit models help mitigate this limitation, and we explored such models with similar results. However, these probit results are much more complex to interpret meaningfully, so we present only the OLS results here. These results are shown in Table 6. The configuration of the independent variables (including controls) is the same as that for the logistic models discussed previously.

Table 6- Relationship Between Income and Perceived Health Index

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	-2.1313 (.0882)**	-2.4680 (.0794)**	-2.3609 (.0825)**
100-133	-2.0088 (.1066)**	-2.1298 (.1024)**	-2.0354 (.1045)**
133-200	-1.4034 (.0782)**	-1.6011 (.0764)**	-1.5235 (.0781)**
200-300	-.9021 (.0692)**	-1.2022 (.0676)**	-1.1465 (.0687)**
300-400	-.7480 (.0746)**	-.9393 (.0725)**	-.9081 (.0726)**
400-500	-.3964 (.0768)**	-.6103 (.0745)**	-.5895 (.0747)**
500-600	-.2042 (.0856)*	-.3451 (.0836)**	-.3318 (.0837)**
Constant	8.991 (.0457)**	10.4248 (.0675)**	10.4131 (.0729)**
R^2	.0647	.1504	.1529
Mean of dependent variable	8.4427	8.4427	8.4427

*Significant at 5% level; **Significant at 1% level.

Table 6 again shows diminishing health returns associated with membership in higher-income cohorts, and the additional information in the data leads to statistically significant (but small in magnitude) negative coefficients for the 500-600% FPL range in each case. This finding indicates that membership in 500-600% FPL group is indeed associated with slightly worse perceived health relative to the reference 600+% FPL group. As we observed in the analysis on weak health, there are also strong marginal health gains associated with membership in "wealthier" groups near the poverty line. Age continues to exert the strongest influence among the controls, although being male is also associated with a .1201 increase in the perceived health index at the 1% significance level.

3.2. Income and Actual Health

A major advantage of using perceived health to measure health status is that a respondent can incorporate all known aspects of health when making his or her evaluation. However, perceived health can be subject to individual biases and inconsistencies, and the subjectivity of the assessment is clearly a drawback of its use. In order to provide a more objective measure of *actual* health, in this section we use the MEPS "priority conditions" in

Table 1 to measure health status. These conditions are much more narrow in scope than "perceived health," but are common enough to provide a rich dataset for analysis. To help give an indication of the individual's overall health, we broaden the metric by creating a *count* of the number of priority conditions with which an individual has been diagnosed. We count all heart-related conditions for a given individual as a single condition in order to prevent over-emphasizing related conditions. Therefore, the maximum count of diagnosed conditions is 8.

For this analysis, we again use a regression model similar to those used previously, with the count of diagnosed conditions as the dependent variable and the same sets of controls used in Section 3.1. However, because we are modeling count data, OLS regression is no longer the most appropriate approach. Instead, the most common alternatives are Poisson regression and negative binomial regression. We explored each of these alternatives, and found the negative binomial regression to be a better fit for the data. Poisson models inherently assume that the mean and standard deviation of the data are equal. In this case, the standard deviation is approximately 1.3 times the mean, indicating the possibility for overdispersion. Although overdispersion is not extreme in this case, a statistically significant alpha (overdispersion coefficient) of .122 resulting from our negative binomial regression with full controls also indicates that this form of model is most appropriate. In practice, the results obtained using each model are similar. The results of the negative binomial regression are shown in Table 7. Coefficients are interpreted as the associated change in the number of conditions as income varies from the 600+% FPL category. Because we use a Poisson regression with sample weights, no value for R^2 is reported.

Table 7- Relationship Between Income and Priority Condition Diagnosis

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	.1171 (.0343)**	.3490 (.0295)**	.3608 (.0308)**
100-133	.2350 (.0453)**	.2147 (.0373)**	.2308 (.0379)**
133-200	.0781 (.0367)*	.1595 (.0304)**	.1726 (.0309)**
200-300	-.0646 (.0336)	.1395 (.0277)**	.1479 (.0280)**
300-400	-.0066 (.0358)	.1400 (.0296)**	.1431 (.0297)**
400-500	-.0827 (.0388)*	.0915 (.0320)**	.0921 (.0320)**
500-600	-.1176 (.0430)**	.0055 (.0364)	.0065 (.0365)
Constant	.0952 (.0217)**	-1.7372 (.0568)**	-1.6867 (.0578)**
Mean of dependent variable	1.1065	1.1065	1.1065

*Significant at 5% level; **Significant at 1% level.

In this case, the impact of age as a control is of much greater magnitude than in the perceived health analyses. In the perceived health analyses, the coefficients for age and income were roughly of the same order of magnitude. For example, in the health index analysis the largest age coefficient was only about 12% greater in magnitude than the largest income coefficient in the model with full controls. In this model, the largest age coefficient (2.5561 for the oldest age group) in the full-control model is over seven times greater than the largest income coefficient. This result is intuitive, as the perceived health questions explicitly asked respondents to "control" for age in their responses by comparing their health to individuals of similar age, and here we deal with absolute diagnoses.

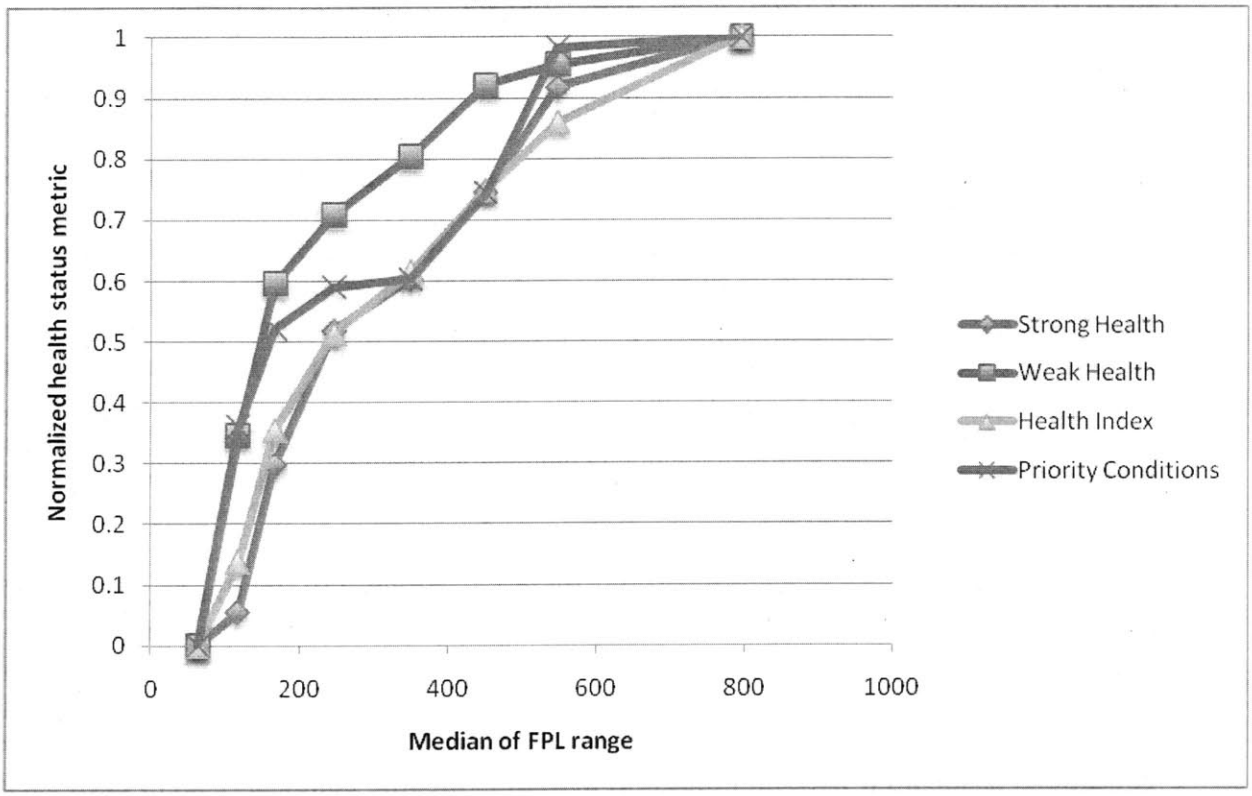
The negative coefficients for income levels above 200% FPL in the "no controls" column are precisely the result of the lack of age controls (since many high income households may also be disproportionately older and sicker by nature). When controls for age are added, the coefficients follow a trend similar to those observed in previous sections, with diminishing decreases to the number of priority conditions associated with membership in higher-income cohorts. Also, in each case with controls, we observe a relatively flat income-health relationship between 200-400% FPL, potentially indicating limited returns to income within the lower middle class. Increased income above that range then is associated with health benefits up to the 500-600% FPL range. The 500-600% FPL coefficient becomes statistically insignificant in both controlled cases, consistent with the diminishing returns to membership in higher-income cohorts we have observed previously.

3.3. Comparison of Income and Health Status Relationships

In the previous two sections, we have compared the magnitude and shape of the income-health relationship in terms of strong perceived health, weak perceived health, a perceived health index, and a count of MEPS priority conditions. While it is

difficult to compare the four sets of metrics directly due to their differences in scale and scope, we feel it is useful to capture the results observed in each of the previous two sections in a single graphic. We linearly normalize each of the four datasets with full controls, with the highest-magnitude coefficient in each dataset mapped to a value of 0, and the omitted 600+% FPL category in each set mapped to a value of 1. The horizontal axis in this case is the median family income in each income cohort as observed in the MEPS dataset. Medians for most cohorts are near the midpoint of the cohot's respective range, with the <100% FPL median at 63% FPL and the 600+% FPL group's median at 795% FPL. Figure 1 shows the results of this comparison.

Figure 1- Comparison of Income-Health Relationships



In Figure 1, we see that the shape of the relationship between income and health status (once normalized) is roughly the same across the four measures of health that we examined. The "weak health" and "health index" curves are the most smooth, while the other two curves exhibit more drastic changes in slope in various regions. The "priority conditions" curve is the most irregularly shaped, likely because it is based on count data and covers a more limited range of health factors than the more subjective measures of perceived health. Assuming causality, the "weak health" and "priority conditions" curves show the strongest returns to income in low ranges (based on their steep slope), again indicating that the marginal returns of income in terms of preventing very bad health (as indicated by perceived

"weak health" or diagnosis with many priority conditions) may be greater than the marginal returns of income in terms of encouraging good health.

3.4. Relationship Between Actual Health and Perceived Health

Our work in the previous three sections shows that low income is associated with both poorer perceived health and poorer actual health. However, it is also important to understand the relationship between actual health and perceived health, and how income impacts this relationship. This understanding will be helpful for interpreting our perceived health results in a broader context, as well as analyzing other data sets where perceived health data might be the only health status information that is available.

To further our understanding of how income affects the relationship between perceived health and actual health, we perform an OLS regression with our "perceived health index" as the dependent variable. As discussed previously, the health index captures all of the perceived health information captured by MEPS into a single dependent variable. Our independent variables include *all* available MEPS health status information from

Table 1 in order to capture as much quantitative information as possible on actual health. We include each of the 11 MEPS "priority conditions" as a separate indicator, rather than relying on count data as we chose to do when using actual health as a dependent variable. We also add four indicators capturing whether the individual needs help with "instrumental" daily activities (e.g., using the phone and cooking) or "other" daily activities (e.g., bathing and dressing), has functional/activity limitations (e.g., walking and bending), or uses assistive technology (e.g., a walker). These additional indicators help broaden the set of controls beyond the priority conditions while still presenting a fairly objective measure of health status. We include the same income independent variables as previously used, as well as controls for age and race.

Table 8 shows the results of our regression relating actual health and perceived health. Similarly to the coefficients shown in Table 6 in Section 3.1, the coefficients shown in the first data column of Table 8 represent the associated change in the perceived health index for a member of the category of interest, relative to the omitted category (i.e., 600+% FPL). A change of magnitude 1.0 represents a change of one response level in one round of the survey. Because we are most interested in the effects of income on the relationship, we focus on income coefficients in the first data column of Table 8. In the second data column, we relate the data obtained here to the data obtained when analyzing the relationship between income and perceived health index without actual health controls in Section 3.1. This calculation is explained in more detail below.

Table 8- Relationship Between Actual Health and Perceived Health

Income Coefficient (% FPL)	Full Controls	Maximum % Difference Unexplained by Actual Health
<100	-1.5655 (.0715)**	66.31%
100-133	-1.4408 (.0937)**	70.79%
133-200	-1.0542 (.0686)**	69.20%
200-300	-.7911 (.0598)**	69.00%
300-400	-.6395 (.0632)**	70.42%
400-500	-.4265 (.0654)**	72.35%
500-600	-.2148 (.0740)**	64.74%
Constant	10.3673 (.0670)**	-
R^2	.3709	-

*Significant at 5% level; **Significant at 1% level.

The R^2 for a perceived health/actual health regression with no controls at all is .3169 (vs. .3709 with full controls here), indicating that much of an individual's perceived health can indeed be explained by actual health, but income and other factors also make a non-trivial contribution to perceived health. Even after controlling for actual health, the data again shows perceived health decreasing with decreasing income, along with diminishing returns to increased income. To put the values into context, coefficients for the 15 actual health variables range from -.1791 for "other" types of heart disease, to -1.7716 for functional activity limitations. The sum of the actual health coefficients (the maximum contribution of actual health to the perceived health index in our model) is approximately -11.5. However, even very sick people are likely only to exhibit only a handful of the 15 conditions, making the income coefficients ranging from -.2148 to -1.5655 quite relevant as predictors of perceived health. For example, a person with income below the poverty line might "under report" his or her perceived health by up to 1.5655 levels over the course of the three MEPS rounds. In the context of the full-controls data in Table 6 relating income and perceived health, up to 66% of the perceived poorer health of a person with income <100% FPL relative to a person with income with income 600+% FPL might be explained simply by the poorer person being more pessimistic about his or her own health[2]. This calculation is shown in the last data column of Table 8 for each income range. However, it is impossible to directly capture all measures of actual health in a model, and the presence of other indicators of actual health would reduce the percentage explained simply by pessimism. Therefore, we can state that the percentages shown in the table are

[2] The income coefficient when we include actual health in the Table 8 regression (decrease in perceived health for a poorer person unexplained by actual health) divided by the income coefficient when we do not control for actual health in the Table 6 regression (total decrease in perceived health for a poorer person) is 1.5655/2.3609 = 66%.

likely the *maximum* percentage of the reduced health that can be explained by pessimism, and in reality, most of the perceived health differences for low-income individuals are likely due to differences in actual health.

4. Income and Healthcare Consumption

Now that we have established an understanding of the income-health status relationship, we seek to answer our second question: Are poorer people likely to consume more or less care than a richer person with the same condition or general level of health? To establish the context of the income-spending relationship, we first examine this relationship using healthcare expenditures with and without a standard set of controls, much as we did for income and health status in Section 3. We then explore the relationship in more detail by adding controls for health status to attempt to quantify the association between income and health spending for *individuals with similar relative health.*

4.1. Income and Health Expenditures

As described in Section 2.2, MEPS provides a variable describing total expenditures for healthcare services for an individual over the course of the year. The expenditures account for payments to providers both by the individual and by other payers, such as health insurers. To examine the relationship between income and health expenditures, we again use an OLS regression similar to those used previously, with MEPS HC-113 2007 healthcare expenditures as the dependent variable. Because possessing health insurance is associated with higher healthcare expenditures and is an easily observable characteristic in practice, we include 2007 health insurance status in our set of possible controls. Table 9 shows the results of this regression. We first report results with no controls, then controlled only for age, and then with a full set of controls that includes all controls from Section 3 in addition to an insurance indicator reflecting whether the individual was covered by any health insurance at any point in 2007. The coefficients of this regression represent the U.S. dollar change in yearly spending associated with membership in each dependent variable category.

Table 9- Relationship Between Income and Health Expenditures

Income Coefficient (% FPL)	No Controls	Controlled for Age	Full Controls
<100	3136.05 (840.93)**	4240.77 (847.40)**	5213.96 (942.78)**
100-133	3628.11 (1201.73)**	2629.15 (1190.36)*	3610.29 (1239.59)**
133-200	2541.10 (816.01)**	2584.26 (777.83)**	3462.80 (804.03)**
200-300	-498.68 (633.19)	544.41 (627.53)	1192.13 (649.88)
300-400	-278.09 (672.65)	541.76 (660.52)	754.12 (664.47)
400-500	-1309.97 (599.21)	-214.58 (586.33)	-73.82 (592.90)
500-600	-685.37 (690.44)	60.25 (673.83)	33.05 (672.71)
Constant	7032.70 (412.64)	1361.61 (452.45)**	-1175.90 (709.15)
R^2	.0035	.0420	.0457
Mean of dependent variable	7514.54	7514.54	7514.54

*Significant at 5% level; **Significant at 1% level.

As we might expect based on the results we observed in Section 3, low incomes are indeed associated with larger healthcare expenditures, and we see diminishing returns (in terms of reduced spending) to health expenditures associated with membership in a higher-income group. Regardless of which set of controls we choose, we observe a drastic drop off in healthcare spending once income reaches the 200-300% FPL range. At and above this income range, healthcare spending is not statistically significantly different from the spending for individuals with family incomes 600% FPL and above. This phenomenon can be explained intuitively because we know that individuals in poor health consume drastically more healthcare than other individuals, and we know the individuals with the lowest family incomes are most likely to be in poor health based on our work in Section 3. In other words, we are seeing compounding curvilinear impacts in the low income ranges, with low income individuals being drastically sicker, and sicker individuals spending drastically more on care than healthier individuals. Among the controls, age continues to possess the strongest association with spending, with a maximum coefficient of 16826.93 for the oldest age group in the full controls case. However, insurance status also exerts a strong influence, and being insured at some point in 2007 is associated with an increase in spending of $3931.23 (also significant at the 1% level) compared to being uninsured throughout 2007.

4.2. Income and Health Expenditures with Health Status Controls

While higher income is indeed associated with lower health expenditures up to a threshold, we know that a large part of this effect is due to the association between income and health. In some cases, it may also be useful to know the association between income and healthcare expenditures *for individuals of similar health status.*

For example, insurers or employers may wish to know if individuals with a certain income level are predisposed towards larger or smaller healthcare spending, regardless of whether they are sick or healthy.

In this section, we expand upon our previous approach by taking measures to control our OLS income-expenditure regressions for health status. As discussed earlier, health status is a very subjective measure, and there are several possible sets of controls available within the MEPS HC-113 data. We start with a set of "baseline controls," which is actually the set of "full controls" used in the previous section, and then we add several additional sets of controls for health status in three separate regressions. The results are shown in Table 10.

The first column of results in Table 10 adds controls for perceived health to the set of baseline controls. Because health status is an independent variable, we are not constrained to our derived metrics of health status. Instead, we can use the individual observations for each of the three MEPS rounds as indicators. Because there are five possible ratings in each round, there are 15 total data points available. We use "excellent" health as the omitted variable for each of the three rounds, leaving 12 total health status indicators in the regression. The second results column of Table 10 uses the 11 MEPS "priority conditions" to control for health status, as well as the four indicators controlling for help with "instrumental" daily activities, "other" daily activities, functional/activity limitations, or assistive technology use. These additional indicators help broaden the set of controls beyond the priority conditions while still presenting a fairly objective measure of health status. We call this set of controls the "conditions controls" set. See

Table 1 for a complete description of these indicators. Finally, the third results column of Table 10 adds both the perceived health controls and the conditions controls to the baseline set.

Table 10- Relationship Between Income and Health Expenditures with Health Status Controls

Income Coefficient (% FPL)	Baseline Controls and Perceived Health Controls	Baseline Controls and Conditions Controls	Baseline Controls, Conditions Controls, and Perceived Health Controls
<100	-635.87 (986.52)	933.54 (964.40)	-1148.36 (994.28)
100-133	-886.04 (1182.24)	183.44 (1134.37)	-1420.46 (1132.38)
133-200	296.71 (780.11)	697.11 (767.85)	-278.34 (763.74)
200-300	-1093.04 (623.54)	-608.39 (612.98)	-1341.61 (613.41)*
300-400	-888.11 (651.84)	-531.07 (646.09)	-1095.11 (642.56)
400-500	-1048.05 (570.51)	-874.00 (556.52)	-1196.69 (554.89)*
500-600	-495.86 (649.70)	-563.03 (647.09)	-659.01 (639.38)
Constant	-623.32 (677.31)	389.40 (687.77)	129.85 (672.44)
R^2	.1128	.1180	.1401
Mean of dependent variable	7514.54	7514.54	7514.54

*Significant at 5% level; **Significant at 1% level.

When only adding controls for perceived health, the association between income and health expenditures is unclear. Most coefficients are negative, indicating that membership in the 600+% FPL cohort may be associated with increased health spending compared to membership in lower-income groups when controlled for health status, but none of the coefficients are statistically significant. The perceived health coefficients dominate the regression, with poor perceived health being associated with an increase in yearly spending of up to $18620.90 (the round three coefficient, significant at the 1% level). Likewise, when only adding the conditions controls, we are unable to see any significant relationship between income and health expenditures. In this case, the highest-magnitude coefficient belongs to the indicator for daily living limitations such as bathing and dressing, and is associated with a $15472.60 increase in yearly spending (significant at the 1% level). Only when we consider all controls for health status do we see any statistically significant relationship between income and health expenditures. In this case, we see that membership in the 200-300% FPL category is associated with a reduction in spending of $1341.61 compared to membership in the omitted 600+% FPL category. Likewise, membership in the "near-wealthy" 400-500% FPL cohort is associated with an $1196.69 decrease in spending relative to the omitted category. All other income coefficients are negative (but statistically insignificant), indicating that the highest earners may have a tendency to consume more healthcare relative to other individuals with roughly the same health status. It is also important to note that the poorest segments of the population *do not* have a tendency to over consume healthcare once we control for health status.

One potential problem with using such a large set of controls is the potential for multicollinearity between the controls, especially those for health status. While there will naturally be some collinearity between them, an analysis of variance inflation factors (VIF) finds no VIF values in the regression above 2.39, far from the threshold of 10 generally accepted as indicating harmful collinearity.

5. Conclusions and Future Work

In this paper, we have examined the relationship between income, health status, and health expenditures for adults using the 2007 data from the MEPS HC-113 household survey conducted by the U.S. Department of Health and Human Services. While this topic has been studied in much detail in the past, we take a unique look at the problem by providing an integrated analysis of income, health status, and health expenditures and answering two key questions regarding the income-health relationship for U.S. adults. First, we determine how much sicker poorer people are than richer people, both in their perception and in actual terms. Second, we determine if a poorer person is likely to consume more or less care than a richer person for given level of health or condition.

We use both perceived health and several "priority conditions" to answer our first question, and in both cases find that low income tends to be associated with reduced health. Consistent with previous research, we find the income/health relationship to be curvilinear, with diminishing returns to health associated with membership in higher-income cohorts. We also find that marginal income at low income levels tends to be more strongly associated with preventing poor health than facilitating strong health, assuming causality in the income-health relationship. We also see that the general shape of the relationship is the same regardless of the health status metric we choose, although in the case of actual health, we saw additional decreased returns to income in the lower middle class. We found a strong relationship between actual health and perceived health across income ranges, although some of our results indicate that poorer people may be more pessimistic about their health than richer people.

To answer our second question, we start by turning our attention to health expenditures. There we find similar results, although the diminishing returns associated with membership in higher-income cohorts are even more drastic than those for health status due to compounding curvilinear effects (between income/health status and health status/health spending). With full controls, we find that the returns to income in terms of association with reduced healthcare expenditures are not significantly different from zero once income reaches the 200-300% FPL range. This finding is in contrast to our health status analyses, where income up to 500-600% FPL was typically associated with improvements in health status. When we add controls for health status to our expenditures models, we find that membership in the highest income group (600+% FPL) is associated with the *highest* level of health spending. This difference is statistically significant at the 5%

level in comparison to the 400-500% FPL "near-wealthy" cohort as well as the 200-300% FPL income group, with expected yearly healthcare spending reductions of over $1100 associated with membership in these groups. As expected, however, the health status effects tend to dwarf income and other effects in the regressions. Our results show that low income is *not* inherently associated with increased healthcare spending once health status is controlled for. Poor people consume more health care because they are sick, not because they are poor. This finding may be encouraging to insurance payers considering participating in the subsidized healthcare exchanges for lower-income households being established as a result of U.S. healthcare reform if they feel they can attract a sufficiently healthy base of patients.

The study of issues related to sociodemographics and health will continue to be an important topic, especially as the U.S. presses forward with healthcare reform, and there are several avenues for future work based on this research. One factor that we do not consider in our analyses due to unavailability of granular MEPS data is geographic location. Much research exists around location issues and healthcare consumption trends and it could be valuable to incorporate location-based controls into this research. Other indicators of healthcare consumption (e.g., hospital discharges) could be explored as well, and a more fine-grained analysis of insurance issues could also be incorporated into our regressions. Finally, the causal nature of the relationships under study could be explored further with a specific eye towards implementation of policies such as income redistribution.

References

CIA. The World Factbook - Fact Listing. September 2010. 28 September 2010 <https://www.cia.gov/library/publications/the-world-factbook/fields/2172.html>.

Der, Geoff, et al. "The Relationship of Household Income to a Range of Health Measures in Three Age Cohorts from the West of Scotland." European Journal of Public Health 9.4 (1999): 271-277.

Gelman, Andrew. "Struggles with Survey Weighting and Regression Modeling." Statistical Science 22.2 (2007): 153-164.

House, James, et al. "Age, Socioeconomic Status, and Health." The Milbank Quarterly 87.9 (1990): 383-411.

Mackenbach, Johan, et al. "The Shape of the Relationship between Income and Self-Assessed Health: An International Study." International Journal of Epidemiology 34.2 (2004): 286-293.

McDonough, Peggy, et al. "Income Dynamics and Adult Mortality in the United States, 1972 through 1989." American Journal of Public Health 87.9 (1997): 1476-1486.

The Henry J. Kaiser Family Foundation. "Focus on Health Reform: Summary of New Health Reform Law." 2010. Kaiser Family Foundation web site. 21 October 2010 <http://www.kff.org/healthreform/upload/8061.pdf>.

U.S Department of Health and Human Services. MEPS HC-113: 2007 Full Year Consolidated Data File. November 2009. 8 October 2010 <http://www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h113/h113doc.shtml>.

U.S. Census Bureau. "The Changing Shape of the Nation's Income Distribution." June 2000. U.S. Census Bureau web site. 28 September 2010 <http://www.census.gov/prod/2000pubs/p60-204.pdf>.

U.S. Department of Health and Human Services. Medical Expenditure Panel Survey. October 2010. 5 October 2010 <http://www.meps.ahrq.gov/mepsweb/index.jsp>.

Wagstaff, Adam and Eddy van Doorslaer. "Income Inequality and Health: What Does the Literature Tell Us?" Annual Review of Public Health 21 (2000): 543-567.