Multidimensional Health Capital and the Production of Health

Tian Liu* Nirav Mehta† Seth Richards-Shubik‡

October 17, 2024

PRELIMINARY—PLEASE DO NOT CITE

Abstract

Health is a complex and multifaceted concept, but to date most empirical research on health capital over the lifecycle models health as unidimensional and often relies on a single observable variable such as self-reported health. In this paper, we develop a dynamic multidimensional factor model for health capital that represents distinct features of health and allows for measurement error in observed proxies for health. We focus on two broad, and distinct, dimensions of health: acute and chronic conditions. We then estimate this model using extremely rich data on health measures and experimental variation in medical input prices from the Rand Health Insurance Experiment (HIE). Our findings indicate that medical care has a stronger effect on chronic health. We also find evidence of unobserved heterogeneity in health production that is correlated with the choice of medical inputs, which would yield a downward bias in the estimated effects of those inputs.

*University of Western Ontario
†University of Western Ontario
‡Lehigh University and NBER

1 Introduction

In a foundational article, Grossman (1972) defined the concept of health capital and showed its importance for models of lifecycle behavior. Much like human capital, introduced by Becker (1962) and Ben-Porath (1967), health capital is a stock that affects the productivity of labor and that can be increased with endogenous investments. Health capital has additional features, in that it may affect utility directly, it determines longevity, and it is subject to risk from substantial negative shocks.

For human capital, a substantial econometric and empirical literature has recently developed to estimate the production technology (e.g., Cunha and Heckman, 2008; Cunha et al., 2010; Agostinelli and Wiswall, 2023; Bono et al., 2020). These papers use dynamic latent factor models to model the evolution of a multidimensional capital stock over time, where the stock itself is not observed but multiple measures of its various dimensions are available. The stocks in a given period depends on the stocks and inputs in the previous period; estimation must take into account that stocks and inputs may all be measured with error. This literature has focused on two dimensions of human capital, cognitive and noncognitive skills, and has shown that human capital is self-productive and indeed multidimensional, and that measurement error is quantitatively important.

By contrast, the literature on health capital over the lifecycle typically models health as unidimensional (following Grossman, 1972) and often relies on a single observable variable, such as self-reported health or number of functional limitations, to serve directly as a perfect measure of the health capital stock. Perhaps due to the fact that, in Grossman (1972)'s terminology, health may yield both "consumption" and "investment" benefits (i.e., direct effects on utility, vs. effects on productivity and longevity), there is not a standard way to define and operationalize health in these models, leaving researchers to choose a particular variable based on the specifics of their empirical application, and, therefore, to focus on a particular aspect of health (see, e.g., Wagstaff, 1986; Gilleskie, 1998; Bolt, 2021; McGee, 2021).[1] At the same time, there are often many measures related to health in commonly used datasets, so this approach necessarily rules out many aspects of health, and further assumes that one aspect is perfectly observable. As a consequence, these models could mismeasure the demand for health and the effects of health on behaviors including medical care utilization and labor force participation.

In this paper, we formulate and estimate a multidimensional dynamic latent factor model for the production of health capital. We focus on two broad, and distinct, dimensions of health: chronic conditions (which have long durations and typically require ongoing treatment) and acute conditions (which onset abruptly but often resolve or can be fully treated within a short time). The effects of investments and shocks on these two dimensions of the capital stock are qualitatively different, as are their effects on utility, productivity, and longevity.

We use data from the Rand Health Insurance Experiment (HIE) (Manning et al., 1987), conducted from 1974 to 1982, to estimate the production technology. The HIE provides numerous measures of chronic and acute conditions over time, as well as exogenous variation in the price of medical care based on random assignment to different health insurance plans. The rich set of health measures and the exogenous variation in prices enable us to apply a dynamic factor model with multiple dimensions of the latent capital stock and with correlated unobserved heterogeneity in the evolution of health and the productivity of medical care.

There are several important benefits from using a latent factor model with multiple dimensions to represent the production and evolution of health in a lifecycle model. First, the factor model makes a distinction between measurement error and true shocks to health capital. This distinction is crucial when modeling risk-averse agents, because only the true shocks affect utility, and hence only they determine the dynamic uncertainty that influences the value of investments in health and of insurance against shocks. If health capital and inputs are only noisily measured or if health capital is truly multidimensional, then models using a particular observed variable as a perfect measure of the health

[1] However see Khwaja (2010) and Cronin (2019) for research using more general models for health.

stock are misspecified. Addressing this misspecification would likely improve estimates of the production technology and would also improve estimates of all parameters in a lifecycle model where health is a central component. Moreover this approach enables researchers to take advantage of the many health measures available in commonly used datasets, without needing to restrict to one variable. This is important in light of the myriad ways in which what we think of as "health" may affect peoples' lifetime utilities.

Second, the multidimensional framework allows for more flexible shock processes than those typically considered in the literature. For example, if there are acute and chronic dimensions to the health capital stock, we may expect the shock processes to differ between the dimensions. Such a framework may be able to capture nuanced dynamics that could be missed by a unidimensional model of health capital, while maintaining the tractable Markov assumptions typically employed by researchers. For example, people with different combinations of acute and chronic health could have the same unidimensional index of health capital stock but have entirely different dynamics due to a change in inputs.

Third, having good estimates of the production technology is particularly important for optimal policy related to health investments, such as health insurance design and prices for medical care services. For example, designing better social insurance would require us to understand the extent of ex-ante uncertainty faced by consumers, as well as the effect of insurance on formal medical care and the effects of those inputs on one's utility, which depends in part on the health production function.

In this preliminary version, we show that the categorization of health into chronic and acute dimensions works well and is supported by the data, and we find that medical care has a stronger effect on chronic health. We also find evidence of unobserved heterogeneity in health production that is correlated with the choice of medical inputs, which would yield a downward bias in the estimated effects of those inputs. To provide a quantitative illustration of the estimated model, we plot the evolution of acute and chronic health for a representative individual under different scenarios. This shows that acute health depreciates more slowly than chronic health, but chronic health responds more to medical care inputs, and that when the price of medical care is lower, more inputs are chosen so chronic health is relatively greater.

In the paper below, we first specify the model (Section 2) and measurement system (Section 3), then describe the data and measure construction (Section 4) and assess the measurement system, where we see that using acute and chronic dimensions works well (Section 4.3). The empirical implementation is described very briefly in Section 5, and the estimates of the model and measurement system are presented in Section 6.

Here we briefly discuss the most relevant literature. This paper is close to the literature specifying and estimating dynamic models of health, where health production plays a key role. In the first structural estimation of a model based on the work of Grossman (1972), Gilleskie (1998) used a dynamic model of acute health to study the determinants and effects of medical care on the duration of sick spells. Khwaja (2010) used a lifecycle model of choice of insurance on medical care utilization and subsequent health outcomes (self-reported health or death) to study willingness to pay for Medicare and Fout and Gilleskie (2015) used a dynamic model to study the choices of health insurance and medical and nonmedical inputs, and subsequent health outcomes, for diabetes patients. More recently, Cronin (2019) used a within-year dynamic model, to study how people choose a health insurance plan and a medical input; notably, this paper also models both acute and chronic dimensions of health. Finally, Bolt (2021) develops and estimates a dynastic model of lifecycle health and cognitive skill investments from childhood through adulthood, focusing on obesity. All of these papers feature a self-productive, or dynamic, health production function, yet none treat the health argument or inputs of the production function as potentially mismeasured.[2]

[2]There is work studying the effect of early resources and investments, namely, education, on later health outcomes (see, e.g., Conti and Heckman, 2010). We allow for education to affect the initial distribution of one's health but then model specific market medical inputs using the detailed data from the HIE. Also somewhat related is research assessing which

Eslami and Karimi (2018) also use the Rand HIE to estimate a production technology mapping determinants of health capital and a medical input (spending), with an interaction, to future health. They do this to help achieve their goal of understanding the relationship between spending on health care and income, both in the cross section and over time. Health is unidimensional (indeed, it needs to be globally ordered for their research goals) and not self-productive in their model (their measures of initial and final health are not comparable, nor do they need to be, for their analysis). Given the quite different goals and focuses of Eslami and Karimi (2018) and our paper, it is most natural to view the two papers as highly complementary.

2 Model

This section presents our model of health capital production, along with a nonstructural approximation of the policy function for medical input choices. Time is discrete. We present the model for one person i, so we suppress that index in this section.

Initial conditions The initial conditions at the start of the Rand HIE can be split into observed variables and unobserved latent factors. Let x_0 measure relevant observable individual characteristics at the time of the start of the HIE (e.g., age at enrollment). Just before the start of the experiment, the health insurance plan p is randomly assigned. There are four continuous latent factors: two dimensions of health capital, as well as permanent income and a latent investment in health (the medical input) described further below. Health capital is comprised of acute and chronic dimensions, and the stock at the end of period t is denoted $h_t = [a_t, c_t]'$. The latent permanent income is denoted y (as the notation indicates, it is assumed to be constant during the study period).

The joint distribution of health capital and income just before the beginning of the HIE is specified as multivariate normal, as follows:

$$[\ln a_0, \ln c_0, \ln y]' \sim N(\mu_0(x_0), \Sigma_0). \tag{1}$$

We also model a persistent, unobserved determinant of health, ζ, where $\ln \zeta \sim \mathrm{N}(0, 1)$, which through the individual's (unspecified) optimization problem may affect the choice of medical inputs. The ζ is distributed independently from the other latent factors.

Policy function We focus on investments in health that are achieved by formal medical care, which we also sometimes refer to as "utilization". We model this medical input as a latent variable, m_t. The data contain experimental variation in the price of medical inputs due to the randomly assigned plans, which will help to estimate the policy function for medical input choices, given the endogeneity of health. For example, some people were assigned to plans in which doctor's visits were free, while others were assigned to plans in which they had to pay 50% of the cost out of pocket (note the HIE was conducted in the US). While we do not derive our input policy function from an explicit lifecycle structural model, it can be thought of an approximation to a Grossman (1972)-type model (see Appendix A).

Given the frequency of corner solutions for the medical input (i.e., zero formal medical care in a given period), it is important that the policy function accommodates them. As we will use the log of the input when modeling health production, we specify a policy function for $\ln \tilde{m}_t \equiv \ln(m_t + \chi)$, where we normalize $\chi = 1$. Specifically, we adopt a "hurdle" model (Cragg, 1971), which could be motivated by the presence of fixed costs from choosing a positive amount of the input (e.g., traveling to a doctor's office). Accordingly, the policy function is comprised of two components: one for the probability of

health measures "best" explain certain economic variables (Blundell et al., 2023) or using factor models to understand health trajectories (Lange and McKee, 2012; White, 2023). While important, this work it does not endogenize health trajectories by using an economic model of health investments.

having zero as the input choice, and another for positive values. The probability that the input is zero (equivalently, $\ln \tilde{m}_t = 0$) is specified as

$$\Pr\{\ln \tilde{m}_t = 0\} = \Lambda\left(\gamma_0^0 + \gamma_x^0 x_t + \gamma_a^0 \ln a_{t-1} + \gamma_c^0 \ln c_{t-1} + \gamma_y^0 \ln y + \gamma_p^0 p + \gamma_\zeta^0 \ln \zeta\right), \tag{2}$$

where Λ is the logistic CDF with location zero and scale one. The component for positive values of the latent input is specified as

$$\ln \tilde{m}_t^+ = \exp\left(\gamma_0^+ + \gamma_x^+ x_t + \gamma_a^+ \ln a_{t-1} + \gamma_c^+ \ln c_{t-1} + \gamma_y^+ \ln y + \gamma_p^+ p + \gamma_\zeta^+ \ln \zeta + \epsilon_t^m\right), \tag{3}$$

where ϵ_t^m is IID with mean zero.[3] From these expressions, it is evident that we must normalize χ because it cannot be separately identified from γ_0^0 and γ_0^+.[4] The x_t includes age, to capture lifecycle effects, and the $\gamma_a^{\cdot}, \gamma_c^{\cdot}$ terms in eq. (2)-(3) allow current health to affect input demand, which could be driven by either valuation of services from health consumption or the effect of health on marginal utility stemming from non-health consumption (i.e., the "investment" channel), as described by Grossman (1972) and more recently laid out in Gilleskie (2021). We note here that the above policy functions abstract from some aspects of the HIE, such as how the out-of-pocket expenditure within a year is capped at a maximum value that depends on one's pre-HIE household income as well as the assigned plan (see, e.g., Cronin, 2019; Hong and Mommaerts, 2021).[5]

Health production function The stock of health evolves based on its lagged value, h_{t-1}, current medical inputs, $\tilde{m}_t$, and exogenous observed characteristics, x_t (e.g., current age). We use a translog specification because it can well approximate a CES specification (Kmenta, 1967), and nests the CES for the special case of Cobb Douglas. Further, recent research on the identification and estimation of dynamic factor models (Agostinelli and Wiswall, 2023; Del Bono et al., 2022; Freyberger, 2021) notes that the translog production function has desirable properties compared to the CES specification.

The production function is as follows:

$$\begin{aligned}
\ln a_t =& \beta_0^a + \beta_a^a[1 + \beta_{a,age}^a age_t] \ln a_{t-1} + \beta_c^a[1 + \beta_{c,age}^a age_t] \ln c_{t-1} + \\
& \beta_{ac}^a[1 + \beta_{ac,age}^a age_t] \ln a_{t-1} \ln c_{t-1} + \\
& \beta_m^a \ln \tilde{m}_t + \beta_{ma}^a \ln \tilde{m}_t \ln a_{t-1} + \beta_{mc}^a \ln \tilde{m}_t \ln c_{t-1} + \beta_\zeta^a \ln \zeta + \epsilon_t^a \\
\ln c_t =& \beta_0^c + \beta_a^c[1 + \beta_{a,age}^c age_t] \ln a_{t-1} + \beta_c^c[1 + \beta_{c,age}^c age_t] \ln c_{t-1} + \\
& \beta_{ac}^c[1 + \beta_{ac,age}^c age_t] \ln a_{t-1} \ln c_{t-1} + \\
& \beta_m^c \ln \tilde{m}_t + \beta_{ma}^c \ln \tilde{m}_t \ln a_{t-1} + \beta_{mc}^c \ln \tilde{m}_t \ln c_{t-1} + \beta_\zeta^c \ln \zeta + \epsilon_t^c.
\end{aligned} \tag{4}$$

The variable ζ represents a persistent, potentially unobservable determinant of health. It would be natural to expect β_ζ^a and β_ζ^c to have the same sign.[6] The structural health shocks $\epsilon_t^a, \epsilon_t^c$ are independently distributed from the latent factors and from each other. The above specification allows for a person's age to affect the evolution of health by affecting the depreciation of the capital stock (e.g., $\beta_{a,age}^a < 0$) and also allows the effect of medical inputs to depend on one's current health. More generally, we

[3] This shock is degenerate in our current implementation.

[4] If there were more than one type of medical input (such as inpatient vs. outpatient utilization, or health behaviors) we would have a separate policy function for each one. This is distinct from having multiple *measures* of a particular medical input (e.g., outpatient expenditures and number of outpatient visits).

[5] In principle, these would not be difficult to explicitly include. For example, different plans would naturally create different incentives for utilization, depending on the distance between out-of-pocket expenditures accumulated thus far this year and one's maximum dollar expenditure.

[6] When estimating we assume $\beta_\zeta^c \geq 0$, and allow the effect of ζ to be unrestricted in the equation for $\ln a_t$ and the policy function.

could allow the entire production function to depend on one's age.[7] We are not restricted to this above specification; e.g., we could add in more polynomial terms.

Our current results are based on a specification that omits the (log) interaction terms:

$$\begin{aligned}\ln a_t =&\beta_0^a + \beta_a^a \ln a_{t-1} + \beta_c^a \ln c_{t-1} + \beta_m^a \ln \tilde{m}_t + \beta_\zeta^a \ln \zeta + \epsilon_t^a \\ \ln c_t =&\beta_0^c + \beta_a^c \ln a_{t-1} + \beta_c^c \ln c_{t-1} + \beta_m^c \ln \tilde{m}_t + \beta_\zeta^c \ln \zeta + \epsilon_t^c.\end{aligned} \tag{5}$$

3 Measurement system

Our model describes the joint distribution of the latent factors, exogenous characteristics (x) and the assigned plan (p): $[[a_0, c_0, y, \zeta]', \{[a_t, c_t, m_t]'\}_{t=1}^T, x_0, \{x_t\}_{t=1}^T, p]'$. The latent factors other than ζ are observed with error as described by a general measurement system, similar to that in Cunha et al. (2010), where we exploit the repeated measures of factors, as suggested by Agostinelli and Wiswall (2023).

We estimate the policy function and health production function using the measures of the medical input and health contained in the HIE data (Manning et al., 1987). The health measures were collected at both enrollment into and exit from the experiment, and the input measures are observed whenever medical care occurred (and are zero otherwise). The data contain a sufficient number of measures to estimate a measurement system for the medical input in all periods as well as health at both enrollment and exit. In addition, experimental variation in health plans (p) provides a source of exogenous variation for medical input choices ($\tilde{m}_t$), which provides exclusion restrictions that allow us to account for the econometric endogeneity of $\tilde{m}_t$ arising from the unobserved heterogeneity ζ in (2)-(3) and (5).

Kotlarski (1967) proves that two independent measures of each latent factor are sufficient to identify the joint distribution of latent factors.[8] Therefore, in this draft we focus on the case in which there are two, "dedicated" measures of each latent factor. However, the latent factors have no inherent scale or location, which makes it difficult to identify the structural equations, as all of the measurement and structural equations are generally permitted to be time-varying functions. See the recent literature studying identification of dynamic latent factor models, Agostinelli and Wiswall (2023); Del Bono et al. (2022); Freyberger (2021). However, Agostinelli and Wiswall (2023) note that there are some instances where the mapping of a latent factor to its measure may be reasonably assumed to be time-invariant. For example, as that paper's application was academic achievement, they used the number of correct responses on the same standardized test administered to students at different ages.

In our application, many measures of health, expenditures, and income would all quite naturally satisfy time/age-invariance. For example, health may decline as one ages, but it is not obvious that the relationship between health and measures of health, such as the presence of chronic conditions or the number of days spent in bed in the previous month[9], should change over time. We note here that our paper's starting point of modeling health as a multidimensional latent variable helps justify this assumption, as we are capturing aspects of health that unidimensional models may not be able to capture, and would therefore likely affect the mapping from such a (restrictive) latent representation of health to measures of health.

To specify the measurement system, let $z_t^{k,j}$ denote the jth measure of the latent factor k in period t (where $k = a, c$), and let $J_t^k \geq 2$ denote the number of measures of k in period t. For each dimension of health $k = a, c$, for the first two measures ($j = 1, 2$) we have:

$$z_t^{k,j} = \lambda_0^{k,j} + \lambda_k^{k,j} \ln k_t + \eta_t^{k,j}, \tag{6}$$

[7] This is because we have time-invariant measures of latent factors; see discussion below.

[8] This result has been generalized in various ways (see, e.g., Hu and Schennach, 2008), but we will not need to use those results here, given the incredibly rich set of measures available in the data.

[9] See Appendix B for a description of this variable and other limitations data we could use.

i.e., the first two measures of a health factor are each a linear function dedicated to only that factor. Additional linear measures follow

$$z_t^{u,j} = \lambda_0^{u,j} + \lambda_a^{u,j} \ln a_t + \lambda_c^{u,j} \ln c_t + \eta_t^{u,j}, \tag{7}$$

i.e., they are undedicated, or may load on both latent health factors. Note that these measures may improve our estimates but are not required for identification of the structural parameters. Similarly, there may also be (i) continuous, nonlinear, measures and (ii) ordinal measures; either may be undedicated.

Our latent income variable is constant over the study period, so we have

$$z^{y,j} = \lambda_0^{y,j} + \lambda_y^{y,j} \ln y + \eta^{y,j}. \tag{8}$$

Measures of m_t may be continuous or ordered, with discrete levels, but in the current version we assume all measures are continuous. The measurement equation for a continuous measure takes into account that it is perfectly measured if there has been no utilization (someone can't be observed to receive market medical inputs if they chose to not use any):

$$z_t^{m,cont,j} = \begin{cases} \lambda_0^{m,cont,j} + \lambda_m^{m,cont,j} \ln \chi & \text{if } \ln \tilde{m}_t \leq \ln \chi \\ \lambda_0^{m,cont,j} + \lambda_m^{m,cont,j} \ln \tilde{m}_t^+ + \eta_t^{m,cont,j} & \text{if } \ln \tilde{m}_t > \ln \chi \end{cases}, \tag{9}$$

where $\eta_t^{m,cont,j}$ is a mean-zero error and, as mentioned above, we normalize $\chi = 1$.[10]

4 Data

We use data from the HIE on males who were at least 18 years old at enrollment,[11] and who were randomly assigned to one of the following plans:

- uniform coinsurance plans, with coinsurance rates of 0%, 25%, 50%, or 95% ("free," "25coin," "50coin," "95coin")
- a plan with 0% / 95% coinsurance for inpatient services / outpatient services ("0in95out")
- a plan with 25% / 50% coinsurance for medical services / dental and outpatient psychiatric services ("25coin50den&men")

(all plans had out-of-pocket limits, see Manning et al. (1987) for details). From this initial sample of 1262, we dropped 143 (11.3%) who were suspended or exited the experiment early. We also dropped individuals with missing data on key variables. The final sample contains 1025 individuals.

From the available data, we use detailed information on health conditions and medical services, as well as demographics and income. The latter come from baseline interviews, and include age, race (whether a person is white or not), years of education, and family income in each of the previous two years. Health information was collected from two sources: self-administered questionnaires and professional medical examinations. These were administered at both enrollment and exit, except that a randomly selected subsample (25-50%, varying by site) was not given medical exams at enrollment. All information on medical services was collected from health insurance claims, which were recorded as part of the experiment.

[10] In principle, this should have a non-negative support (e.g., log-normal), so as to avoid censoring at zero. In practice, however, our estimates of the distributions of both measures of $\tilde{m}$ do not exhibit a large amount of truncation at zero.

[11] Individuals from the Dayton site were excluded because they had different questionnaires at enrollment.

4.1 Measure construction

We use a large number of raw variables on health conditions to construct two summary measures of acute health and two summary measures of chronic health. Each summary measure is generated using a factor analysis (we use the predicted factor score from a single-factor model, estimated via maximum likelihood) of a distinct set of raw variables. For each dimension of health capital, the goal is to have two measures of a similar construct, so when there are multiple variables related to the same condition, we split them between the two measures for that dimension, as detailed below.

For chronic conditions, we use the Medical Disorder Series files, which contain information from both the questionnaires and medical exams, although most of the variables we choose come from the latter. The specific variables used for each measure of chronic health are as follows:

- Chronic Measurement 1: hemoglobin value, sum of abnormality in electrocardiogram, forced expiratory volume in one second, average hearing threshold for right ear, blood glucose, blood cholesterol, systolic blood pressure, serum uric acid, total serum T4, and far vision
- Chronic Measurement 2: mean cell volume, Minnesota Code score from electrocardiogram, average hearing threshold for left ear, hypercholesterolemia (from questionnaires), diastolic blood pressure, blood urea nitrogen, near vision

Acute conditions do not have good coverage in the Medical Disorder Series. Instead we use information from the self-administered questionnaires contained in other files, on health issues such as cough, sore throat, skin rash, and bursitis. The variables come from questions that ask whether individuals had certain symptoms or conditions, or experienced pain or concern about those conditions, over the past 30 days to 12 months. From each raw variable, we make binary variables for any indication of the condition, and the factor analysis is applied to the latter. The specific variables used for each measure of acute health are as follows:

- Acute Measurement 1: cough, vomiting, backaches, bursitis and skin rash
- Acute Measurement 2: cold, hernia, tuberculosis, and sprained ankle

Again, the reason to split the variables between the two measures is to have them relate to the same or similar conditions, such as cough and cold.

After obtaining factor scores, we reverse the factor scores and use those as our measures, so higher scores are "better".

Last, we measure medical inputs with spending on outpatient care and number of outpatient visits (excluding dental care). The raw data record the dates of service, and we use this to compute total expenditures and visits for every period.[12] This yields a consistent longitudinal series on the two measures of the medical input construct (which often equal zero in any given period).

4.2 Summary statistics

We currently consider the case where we have two dedicated measures of each factor. Table 1 summarizes our dedicated measures of each of the latent factors.

Table 2 shows means and standard deviations for important initial characteristics (such as education and age at enrollment) as well as measures of income and health. Table 3 shows the distribution of our sample across assigned plans. The plurality of individuals were assigned to the free care plan, followed by the plan covering all inpatient services but requiring the individual to cover 95% of outpatient services ("0in95out"). As noted above, the plan "25coin50den&men" featured a 25% coinsurance rate for all services excepting dental and mental health, for which there was a 50% coinsurance rate.

[12] Four-week periods are natural to use because the experiment paid participation incentives every four weeks.

Table 1: Dedicated measures of latent factors

Factor	j	Observed at	Measure
a_t	1	enroll/exit	reversed factor score, as described in Section 4.1
	2	enroll/exit	reversed factor score, as described in Section 4.1
c_t	1	enroll/exit	reversed factor score, as described in Section 4.1
	2	enroll/exit	reversed factor score, as described in Section 4.1
$\tilde{m}_t$	1	each period	log of one plus the value of outpatient expenditures
	2	each period	log of one plus the number of outpatient visits
y	1	enroll	log of household income the year before enrollment
	2	enroll	log of household income two years before enrollment

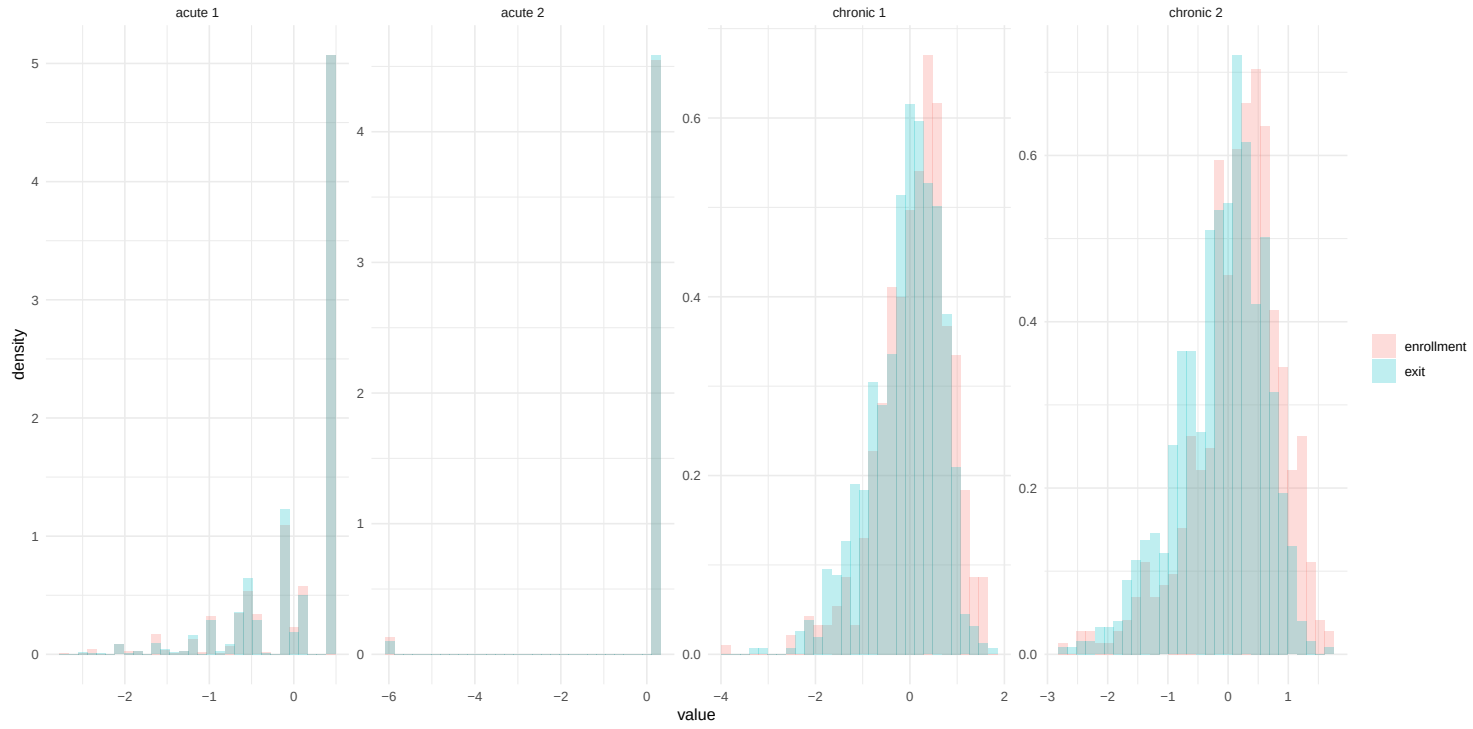
Table 2: Summary statistics for estimation sample

	Mean	SD	N
education years at enro	12.3300	3.2300	1025
white	0.8600	0.3400	1025
age at enro	36.4400	11.0300	1025
log of family income (year preceding enro)	9.1700	0.6400	968
log of family income (2nd year preceding enro)	9.1500	0.7600	650
acute measurement 1 at enro	−0.0080	0.5900	1009
acute measurement 1 at exit	0.0060	0.5600	1008
acute measurement 2 at enro	−0.0100	1.0300	1012
acute measurement 2 at exit	0.0400	0.8800	1008
chronic measurement 1 at enro	0.1100	0.7500	476
chronic measurement 1 at exit	−0.1000	0.7500	811
chronic measurement 2 at enro	0.1500	0.7300	475
chronic measurement 2 at exit	−0.1200	0.7200	810

Table 3: Distribution of sample across assigned plan

plan	N	Percent
free	365	35.6100
25coin	94	9.1700
25coin50den&men	103	10.0500
50coin	41	4.0000
0in95out	265	25.8500
95coin	157	15.3200

Figure 1: Distributions of constructed health measures



4.3 Measure assessment

Here we assess joint distributions of the constructed measures of health capital and medical inputs. To support the estimation of the measurement system and the model, these measures must have sufficient variation individually, as well as correlations among related measures.

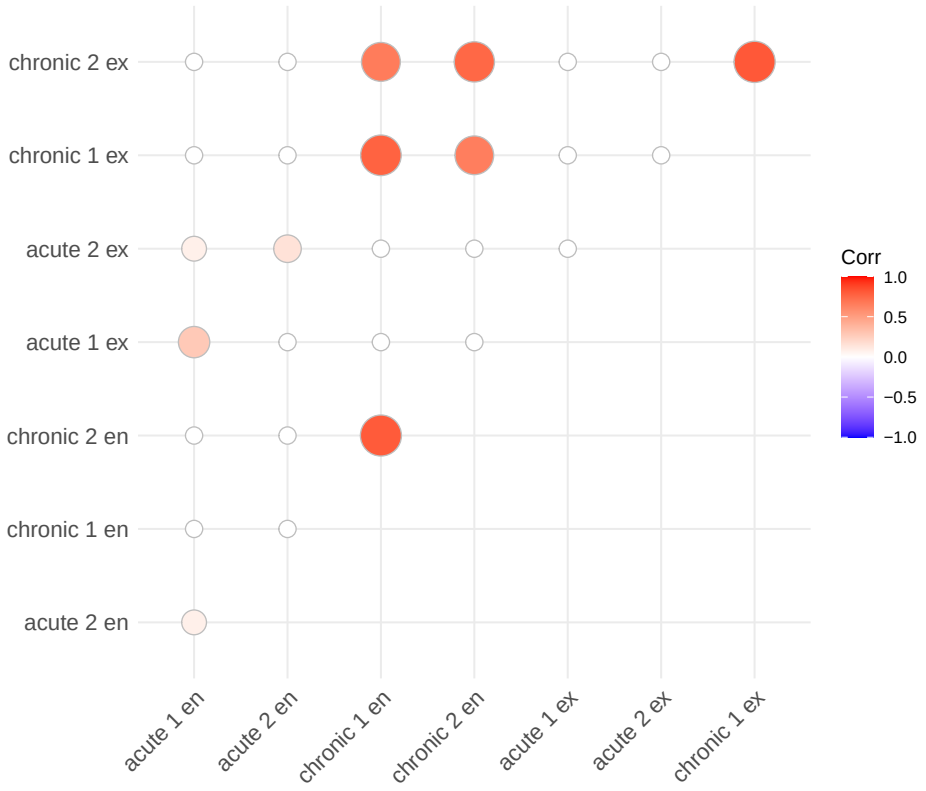
Figure 1 plots the distributions of the health measures at enrollment (pink) and exit (blue). The chronic measures exhibit more variation, with many more points of support, compared to the acute measures, especially the second acute measure ("acute 2"). Also, the chronic measures show some deterioration over time, while the acute measures do not. The distributions of the chronic measures at exit (blue) are somewhat to the left of the distributions at entry (pink), while the distributions of the acute measures largely overlap. This is consistent with ongoing nature of chronic conditions, which may depreciate the stock of chronic health, and the temporary nature of acute conditions.

Figure 2 visualizes the joint correlations among the acute and chronic health measures at entry (en) and exit (ex).[13] The positive and significant contemporaneous correlations between the two acute measures at entry and between the two chronic measures at entry indicate that each pair of measures is indeed capturing a latent dimension of health. Furthermore, the correlations over time are stronger for the chronic measures than for the acute measures, which fits with the basic distinction between chronic and acute conditions. Thus it appears that our two sets of measures are capturing distinct dimensions of health capital, and these dimensions exhibit patterns consistent with their labels as chronic and acute health.

Next we consider the measures of medical inputs. The first and second rows of Table 4 present the mean cost (here and hereafter, this is the total expenditure, and includes both out-of-pocket and insured amounts) and number of visits, per four week period, respectively, split by whether the health insurance plan had zero coinsurance (the "free plan" labeled "Free") or positive coinsurance (labeled

[13]Only the upper triangle is displayed. Correlations that aren't significantly different from zero at the 5% level appear as empty circles.

Figure 2: Correlations among dedicated health measures



"Coins."). As expected, both input measures are significantly higher for people assigned to the free plan. The next two rows show the probabilities of having positive costs (third row) and positive visits, again split by plan type, and show the same pattern: that those assigned to the free plan have a higher likelihood of positive input levels. The last two rows show that, while the average levels of costs and visits are also higher when considering only the positive values, the differences here are smaller than those in the probabilities of positive amounts. That is, the differences in average levels evinced in the top two rows is driven by the probability of positive amounts, not the intensive margin. This is also borne out by the plots in Figure 3, of the distributions of the costs (left panel) and visits (right).

Because the measures of medical inputs systematically vary based on the assigned plan, we can use the plans to examine the possibility of an exogenous relationship between medical inputs and health capital. To do this, we estimate regressions of the health measures at exit on insurance plan characteristics: an indicator for being assigned to the free care plan and the amount of the out-of-pocket limit. The regressions also control for demographics and insurance status prior to the experiment; while not necessary to ascribe a causal interpretation to the coefficients on plan characteristics, these variables may help capture health at the start of the experiment and, as such, help with interpretation, and also they may improve precision by reducing the residual variance. Table 5 presents the results. There is no discernible relationship with the acute measure, but there is an effect on the chronic measure, from both the out-of-pocket limit and being assigned to the free plan. This suggests that medical inputs are mainly relevant for chronic health, while acute health may be affected more by random shocks.

Table 4: Medical input measures

Variable	Not free	Free	p-value[a]
Cost ($)	12	18	<0.001
Visits (#)	0.27	0.40	<0.001
Positive cost	21%	29%	<0.001
Positive visits	15%	22%	<0.001
Cost ($) | positive cost	59	61	<0.001
Visits (#) | positive visits	1.82	1.84	0.061

Main columns show means or percentages of levels (not logs) of cost in dollars or outpatient visits.
[a]Wilcoxon rank sum test or Pearson's Chi-squared test

Figure 3: Distributions of levels of costs and doctors visits

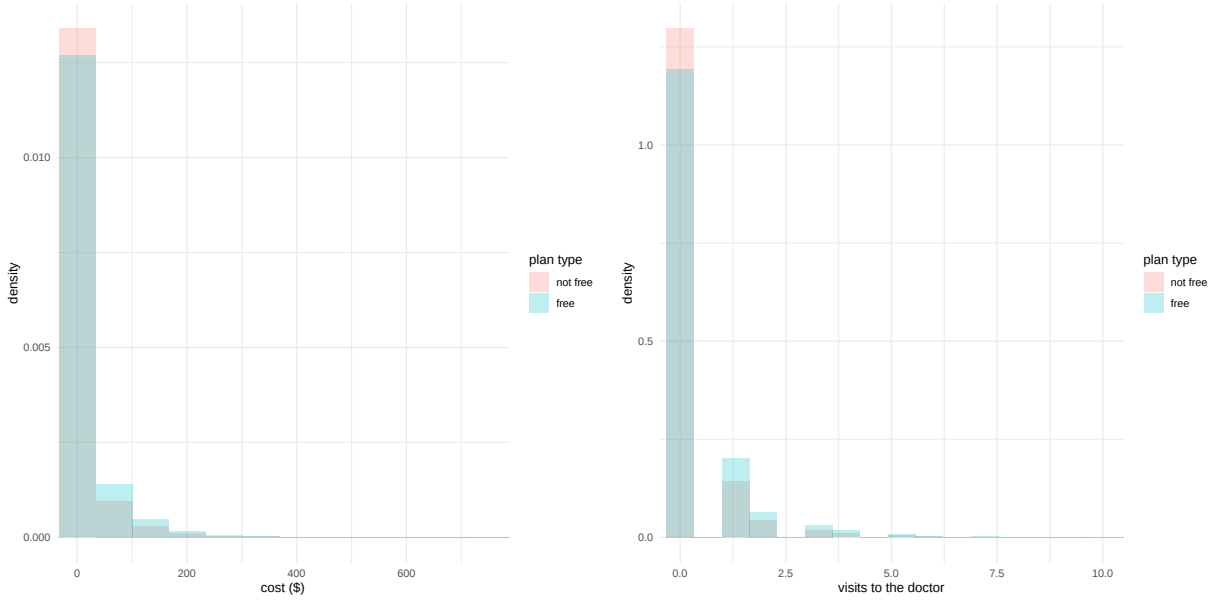


Table 5: Regressions of first health measures at exit on initial conditions and plan

	acute measurement 1 at exit	chronic measurement 1 at exit
	(1)	(2)
age at enro	0.001	-0.040^{***}
	(0.002)	(0.002)
white	-0.051	0.443^{***}
	(0.056)	(0.064)
education years at enro	0.006	0.034^{***}
	(0.006)	(0.007)
only private insurance at enro	0.025	-0.177^{*}
	(0.083)	(0.092)
only work insurance at enro	-0.011	-0.100
	(0.062)	(0.070)
both work and private insurance at enro	-0.036	-0.033
	(0.079)	(0.088)
out-of-pocket limit	-0.00002	0.0003^{***}
	(0.0001)	(0.0001)
free care plan	-0.005	0.132^{*}
	(0.061)	(0.068)
Constant	-0.040	0.517^{***}
	(0.119)	(0.133)
N	967	777
R^2	0.002	0.462
F Statistic	0.278	82.459^{***}

*p $<$ .1; **p $<$ .05; ***p $<$ .01

Table 6: Production function parameters

acute	estimate	chronic	estimate
β_0^a	-0.0012	β_0^c	-0.0163
β_a^a	1.0203	β_a^c	0.0069
β_c^a	0.0003	β_c^c	1.0031
β_m^a	0.0008	β_m^c	0.0112
β_ζ^a	0.0015	β_ζ^c	0.0081
$\ln \sigma^2_{\epsilon_a}$	-21.5831	$\ln \sigma^2_{\epsilon_c}$	-29.9998

5 Empirical Implementation

We assume all of the measurement errors are independently normally distributed. We estimate the model parameters using maximum likelihood, following the the same global algorithm as Cunha and Heckman (2008). We start by using our analytical solutions for the latent factors and measurement system, and estimating the technology and policy functions given these values. We then maximize over all parameters, using those obtained previously as initial guesses.

6 Estimates

Table 6 presents the preliminary point estimates for production function parameters.Consistent with the results presented in Table 5, we can see that the effect of the health input is negligible for acute health (β_m^a), but not so for chronic health (β_m^c).

Table 7 presents the policy function parameters. The first column presents estimates of coefficients governing the probability $\tilde{m} = 0$ and the second column presents estimates governing the behavior of

positive values of $\tilde{m}$. Older people had higher probabilities of positive input choices ($\gamma_x^0 < 0$) and higher input levels conditional on a positive input choice ($\gamma_x^+ > 0$). Higher levels of acute or chronic health reduce the likelihood of positive input levels ($\gamma_a^0, \gamma_c^0 > 0$) but different effects on the input conditional on it being positive.

Table 7: Policy function parameters

Prob($\tilde{m} = 0$)	estimate	$\tilde{m}$ if positive	estimate
γ_0^0	1.6879	γ_0^+	1.2443
γ_x^0	−0.0064	γ_x^+	0.0009
γ_a^0	0.9377	γ_a^+	−0.1176
γ_c^0	0.2029	γ_c^+	0.0560
γ_y^0	−0.0188	γ_y^+	−0.0046
γ_ζ^0	1.1644	γ_ζ^+	−0.0611
Plan indicators (γ_p^0):		Plan indicators (γ_p^+):	
free (base)	0.0000	free	0.0000
25coin	0.1319	25coin	0.0091
25coin50den&men	0.0817	25coin50den&men	0.0176
50coin	0.1842	50coin	−0.0013
0in95out	0.4143	0in95out	−0.0333
95coin	0.6263	95coin	−0.0437

Table 8 presents the parameters governing the distribution of initial latent factors. The top panel presents the means of the latent factors. All else equal, people who were older at entry had higher average incomes, lower average values of chronic health, and, somewhat curiously, higher average values of acute health (though this last effect is small). People with more education and those who are white all have higher average levels of both dimensions of health as well as income. The second panel presents the parameters for the jointly distributed deviates of initial acute and chronic health and income. There is more dispersion in the chronic health deviate than the one for acute health. The deviates for initial chronic and acute health are negatively correlated, but those for acute health and income and chronic health and income are both positively correlated.

Table 9 presents the estimates of the measurement system parameters, and shows the substantial amount of variance in each measurement attributable to measurement error. For example, the variance (pooling over enrollment and exit) of the first measure of acute is 0.33, meaning the measurement error accounts for the vast majority of its variation. In contrast, the pooled variance of the first chronic health is 0.57, which means the measurement error accounts for a smaller share of this health measure's variation.

6.1 Trajectories and comparative statics under estimated parameters

To get a feel for how health evolves we first plot the trajectory of acute and chronic health from enrollment until five years later for a somewhat "typical" example person, who was white, age 36 at enrollment, and a high school graduate. We set the initial conditions to be at their mean values for a person with these characteristics (i.e., we set the deviations of the latent factors from their means to zero, including for $\ln \zeta$).[14] We then simulate forward, using a single random sequence of shocks to

[14] For reference, note that the standard deviations of initial (period-0) acute and chronic health (which are unaffected by plan assignment) are, respectively, 0.219 and 0.718.

Table 8: Initial distribution of (logs of) latent factors

Factor structure			
Means of initial conditions			
acute		chronic	
intercept	−0.3219	intercept	1.0368
coef. on age at enro	0.0028	coef. on age at enro	−0.0406
coef. on educ. years at enro	0.0138	coef. on educ. years at enro	0.0276
coef. on white	0.0487	coef. on white	0.2349
income			
intercept	7.5862		
coef. on age at enro	0.0235		
coef. on educ. years at enro	0.0396		
coef. on white	0.2649		
Distribution of deviation from means			
$Var(\ln a_0 \mid x_0)$	−3.0949	$Cor(\ln a_0, \ln c_0 \mid x_0)$	−0.2114
$Var(\ln c_0 \mid x_0)$	−1.2882	$Cor(\ln a_0, \ln y \mid x_0)$	0.1308
$Var(\ln y \mid x_0)$	−0.9052	$Cor(\ln c_0, \ln y \mid x_0)$	0.0878

Note: x_0 is a vector containing the person's age at enrollment, years of education at enrollment, and an indicator for whether the person is white.

Table 9: Estimates of measurement system

	acute		chronic		income		$\tilde{m} \mid \tilde{m} > 0$	
Parameter	measure 1	measure 2	measure 1	measure 2	measure 1	measure 2	measure 1	measure 2
Intercept (λ_0)	−0.0106	0.0407	0.1110	0.1180	0.0000	−0.0107	0.0000	0.0000
Loading (λ_1)	1.0000	0.2509	1.0000	0.7505	1.0000	1.0000	1.0000	0.2714
Error variance (σ^2_η)	0.2814	1.0218	0.1428	0.2289	0.0266	0.1890	1.1098	0.1424

Note: The loading on the first measure of each factor is normalized to one. Further, because of the natural scale of the input measures (zero versus positive), the intercept of each is set to zero. We also fixed the loading of the second income measure to one and the intercept for the first income measure to zero.

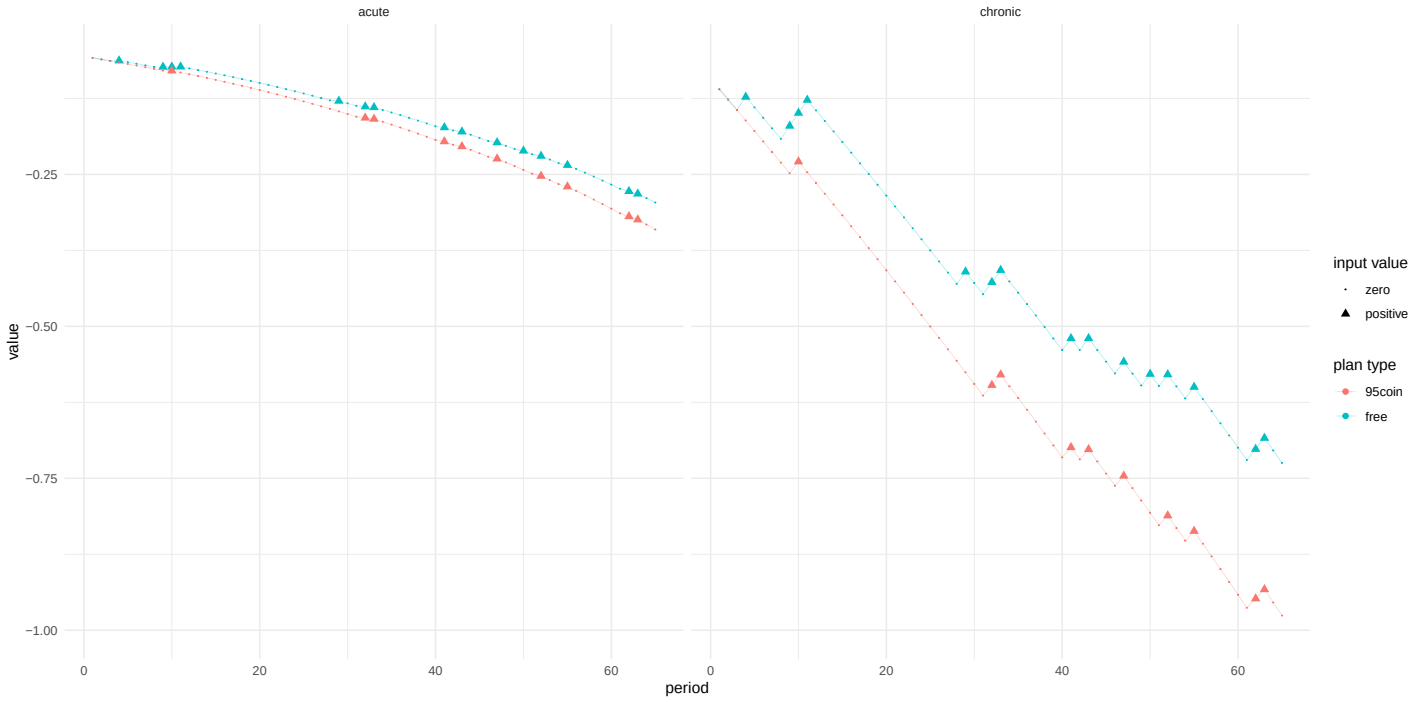
health $(\epsilon^a_t, \epsilon^c_t)$ and medical input choices $(\epsilon^0_t, \epsilon^m_t)$.[15]

Figure 4 plots the trajectories of acute (left panel) and chronic (right panel) health, if this hypothetical person were assigned to the free (blue) or 95% coinsurance (red) plans. The dots correspond to periods in which the input was zero and the triangles correspond to periods in which there was positive medical care utilization; we focus on the extensive margin of utilization here because that was where the most variation lies, according to the input measures. Starting with acute health, the paths under the free and 95% coinsurance plans decline at identical rates until the fourth period, in which there is a positive input level under the free plan but not under the 95% coinsurance plan.[16] The same is true for chronic health, although the decline is steeper. Consistent with the estimates of the effect of health on input choices, there are more positive values of the input in later periods, as both dimensions of

[15]The shocks ϵ^0_t are distributed IID uniformly on the unit interval; periods in which they exceed the probability of the input being zero according to eq. (2) result in positive input amounts, given by eq. (3).

[16]Because we are fixing the initial conditions and the shocks are identical, if there is positive utilization under the 95% coinsurance plan there will be a positive input level under the free plan, because the estimated coefficient on the probability of zero is positive for the 95% coinsurance plan (see Table 7).

Figure 4: Health for example person with $\ln \zeta = 0$, by plan type

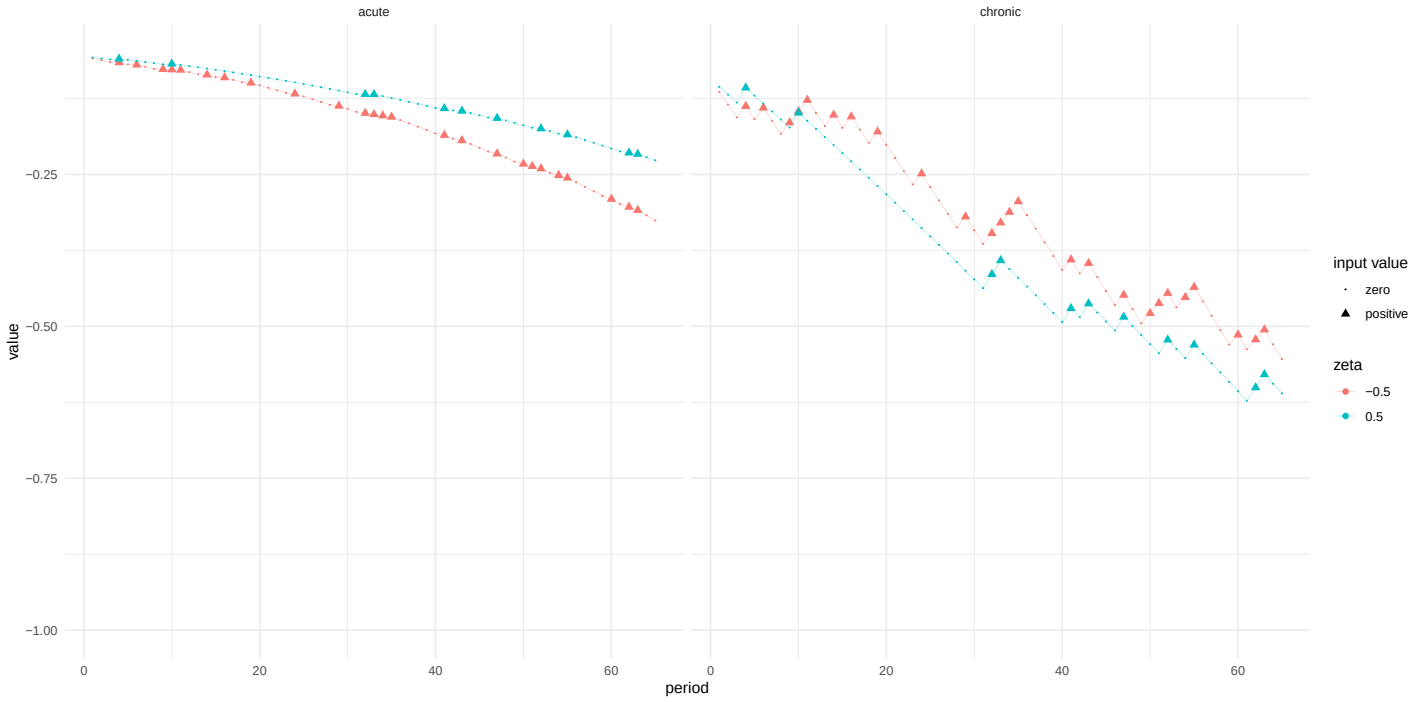


health decline; for example, the input was positive only once under the 95% coinsurance plan during first 20 periods and five times during the last 20 periods. Also, the figure shows the somewhat negligible effect of utilization on acute health and the more substantial effect on chronic health, as seen with the parameter estimates in Table 6.

Next, to understand how the permanent unobserved heterogeneity ζ affects outcomes, Figure 5 plots health trajectories under the free plan, using the identical sequence of shocks as above, for two values of $\ln \zeta$: a low value, half a standard deviation below its mean (red); and a high value, half a standard deviation above its mean (blue). As above, the triangles denote periods in which the medical input was positive. The path of acute health for the low value of $\ln \zeta$ is similar to that under the free plan when $\ln \zeta = 0$, shown in Figure 4. At our estimated parameters, lower values of ζ increase the probability of a positive input but also decrease acute health over time. The path of acute health under the high value of $\ln \zeta$ lies above that for the low value, as the direct effect of the higher value of ζ in the production function for acute health outweighs the indirect effect that operates via the lower input amounts under that value of ζ, due to the small effect of the input on acute health, resulting in a shallower decline in acute health.

Moving on to chronic health, the path for the higher value of ζ lies discernibly above than that for the low value even before the first positive input. This contrasts with the paths of acute health in the same periods, which are closer together due to the smaller coefficient on $\ln \zeta$ in the production of acute health. However, the effect of the lower levels of inputs chosen under the high value of ζ outweighs the otherwise flatter decline in chronic health, stemming from the direct effect via production. The result is that, for this example at least, the higher value for unobserved heterogeneity results in higher acute health and lower chronic health.

Figure 5: Health for example person under free plan, by $\ln \zeta$



References

Agostinelli, F. and M. Wiswall, "Estimating the Technology of Children's Skill Formation," *Journal of Political Economy*, 2023.

Becker, G. S., "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 70(5, Part 2):9–49, 1962.

Ben-Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4):352–365, 1967.

Blundell, R., J. Britton, M. Costa Dias and E. French, "The Impact of Health on Labor Supply near Retirement," *Journal of Human Resources*, 58(1):282–334, 2023.

Bolt, U., "What is the Source of the Health Gradient? The Case of Obesity," *Working paper*, 2021.

Bono, E. D., J. Kinsler and R. Pavan, "A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation," 2020.

Conti, G. and J. J. Heckman, "Understanding the Early Origins of the Education–Health Gradient: A Framework That Can Also Be Applied to Analyze Gene–Environment Interactions," *Perspectives on Psychological Science*, 5(5):585–605, 2010.

Cragg, J. G., "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, pp. 829–844, 1971.

Cronin, C. J., "Insurance-Induced Moral Hazard: A Dynamic Model of Within-Year Medical Care Decision Making under Uncertainty," *International Economic Review*, 60(1):187–218, 2019.

Cunha, F. and J. J. Heckman, "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4):738–782, 2008.

Cunha, F., J. J. Heckman and S. M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3):883–931, 2010.

Del Bono, E., J. Kinsler and R. Pavan, "Identification of Dynamic Latent Factor Models of Skill Formation with Translog Production," *Journal of Applied Econometrics*, 37(6):1256–1265, 2022.

Eslami, K. and S. M. Karimi, "Health Capital and the Productivity of Health Spending," *Working Paper*, 2018.

Fout, B. T. and D. B. Gilleskie, "Does Health Insurance Encourage or Crowd Out Beneficial Nonmedical Care? A Dynamic Analysis of Insurance, Health Inputs, and Health Production," *American Journal of Health Economics*, 2015, doi: 10.1162/AJHE_A_00010.

Freyberger, J., "Normalizations and Misspecification in Skill Formation Models," 2021.

Gilleskie, D. B., "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, pp. 1–45, 1998.

—, "In Sickness and in Health, until Death Do Us Part: A Case for Theory," *Southern Economic Journal*, 87(3):753–768, 2021.

Grossman, M., "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2):223–255, 1972.

Hong, L. and C. Mommaerts, "Time Aggregation in Health Insurance Deductibles," Tech. rep., National Bureau of Economic Research, 2021.

Hu, Y. and S. M. Schennach, "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1):195–216, 2008.

Khwaja, A., "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.

Kmenta, J., "On Estimation of the CES Production Function," *International Economic Review*, 8(2):180–189, 1967.

Kotlarski, I., "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20(1):69–76, 1967.

Lange, F. and D. McKee, "The Evolution of Latent Health Over the Life Course," *Working Paper*, 2012.

Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler and A. Leibowitz, "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review*, pp. 251–277, 1987.

McGee, R., "Old Age Savings and House Price Shocks," *University of Western Ontario Department of Economics Research Reports*, (2021-4), 2021.

Wagstaff, A., "The Demand for Health: Some New Empirical Evidence," *Journal of Health Economics*, 5(3):195–233, 1986.

White, M. N., "Self-Reported Health Status and Latent Health Dynamics," *Journal of Health Economics*, 88:102727, 2023.

A Lifecycle model underlying our policy function

We sketch a Grossman (1972)-type model of investments in health capital, the policy function of which is approximated by our policy function. For simplicity, we present this model without accounting for details of the Rand HIE other than the random assignment of the insurance plans (e.g., we do not explicitly account for the start/end of the Rand HIE here, or within-year effects due to out-of-pocket limits). We also abstract from the sources of uncertainty that render the input choice a stochastic function of the period's pre-determined state variables in our specification.

The agent's period utility function is

$$U_t = U(\phi_t(h_t), g_t), \tag{10}$$

where the function $\phi_t(\cdot)$ converts health capital stocks in period t into service flows and g_t represents utility flows from non-health services.

Let y_t denote exogenous income in period t (in our specification we treat income as fixed, as the HIE data cover only up to five years of the lifecycle) and define $\mathbf{y} \equiv \{y_t\}_{t=1}^T$, and analogously define $\mathbf{x}$. The agent's problem is

$$\max_{\{m_t\}_{t=1}^T} \sum_{t=1}^{T} \beta^{t-1} \mathrm{E}_t \left[U(\phi_t(h_t), g_t)\right] \tag{11}$$

s.t.

$$h_t = K_t(h_{t-1}, m_t; \mathbf{y}, \mathbf{x}, \zeta, \epsilon_t^a, \epsilon_t^c) \tag{12}$$

$$g_t = G_t(h_t, m_t; p, \mathbf{y}, \mathbf{x}, \zeta) \tag{13}$$

$$(h_0, \mathbf{y}, \mathbf{x}) \text{ given}, \tag{14}$$

where K_t is a (vector-valued) health production function and $G_t(\cdot)$ includes h_t as an argument due to the potential investment effect of health, and includes the input m_t to reflect the cost of inputs. The budget constraint is captured in the function $G_t(\cdot)$. The variable p inside G_t measures shifters to the cost of inputs m_t, which are independently distributed from all other variables. The random assignment of people to health insurance plans with different cost-sharing features in the Rand HIE ensures that this independence is satisfied.

The maximization of (11) results in the policy functions

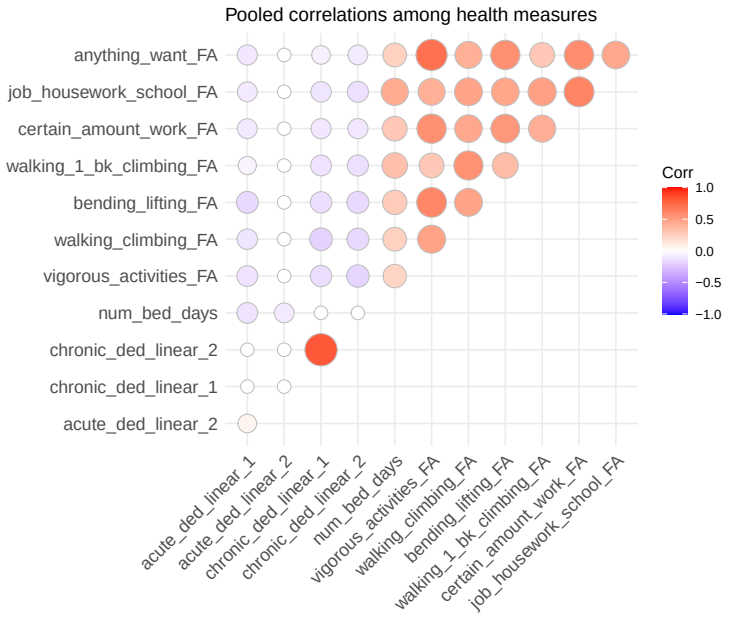
$$m_t^*(h_t, \mathbf{y}, \mathbf{x}, \zeta, p),$$

where the presence of ζ makes explicit the need to address econometric endogeneity when estimating the technology (12).

B Limitations data

Here, we present information about certain undedicated health measures (i.e., those not used in the measurement system) and how these relate to the latent health factors. These measures will be used later as outcome variables, to illustrate features of the model quantitatively. Figure 6 plots the correlations among the dedicated health measures and physical limitations measures also available in the surveys/screening exams, pooling observations at entry and exit. The physical limitations measures are all coded such that higher values reflect more limitations. The first limitations measure records the number of days out of the last 30 days the respondent was confined to bed. The other limitations measures have been converted to binary variables. We can see there's a significant negative correlation between the acute health measures and days in bed, but no such relationship between the chronic health

Figure 6: Pooled correlations among dedicated health and physical limitations measures



measures and days in bed. However, both acute and chronic health measures are significantly negatively correlated with many of the physical limitations measures.

Table 10 presents estimates of the physical limitations measurement system, where the second dedicated measures of acute and chronic have been used to instrument for the first dedicated measures. We can see that many of the limitations measures load on both acute and chronic dimensions of health.

Table 10: Estimates of measurement system for physical limitations

	Dependent variable:							
	num.bed.days	vigorous.activities.FA	walking.climbing.FA	bending.lifting.FA	walking.1.bk.climbing.FA	certain.amount.work.FA	job.housework.school.FA	anything.want.FA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
acute_ded_linear_1	5.274**	0.378	0.232*	0.514**	0.185*	0.401**	0.254*	0.561**
	(2.289)	(0.237)	(0.137)	(0.241)	(0.099)	(0.194)	(0.137)	(0.271)
chronic_ded_linear_1	0.160	0.097***	0.051***	0.069***	0.027***	0.042***	0.038***	0.048***
	(0.144)	(0.016)	(0.009)	(0.016)	(0.006)	(0.013)	(0.009)	(0.018)
Constant	0.219**	0.107***	0.029***	0.071***	0.011***	0.042***	0.023***	0.085***
	(0.089)	(0.010)	(0.006)	(0.010)	(0.004)	(0.008)	(0.006)	(0.011)
Observations	1,357	1,372	1,373	1,373	1,373	1,372	1,373	1,373

Note: *p<0.1; **p<0.05; ***p<0.01