NBER WORKING PAPER SERIES

EARLY CHILDHOOD EDUCATION AND LIFE-CYCLE HEALTH

Jorge Luis García
James J. Heckman

Working Paper 26880
http://www.nber.org/papers/w26880

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2020

The authors wish to thank Frances Campbell, Craig and Sharon Ramey, Margaret Burchinal, Carrie Bynum, and the staff of the Frank Porter Graham Child Development Institute at the University of North Carolina Chapel Hill for the use of data and source materials from the Abecedarian Project and the Carolina Approach to Responsive Education. Collaboration with Bryan Tysinger of the Leonard D. Schaeffer Center for Health Policy and Economics at the University of Southern California on adapting the Future America Model is gratefully acknowledged. James J. Heckman: This research was supported in part by: NIH grants NICHD R37HD065072 and NICHD R01HD054702; and the American Bar Foundation. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Early Childhood Education and Life-cycle Health
Jorge Luis García and James J. Heckman
NBER Working Paper No. 26880
March 2020
JEL No. C93,I10,I28,J13

ABSTRACT

This paper forecasts the life-cycle treatment effects on health of a high-quality early childhood program. Our predictions combine microsimulation using non-experimental data with experimental data from a midlife long-term follow-up. The follow-up incorporated a full epidemiological exam. The program mainly benefits males and significantly reduces the prevalence of heart disease, stroke, cancer, and mortality across the life-cycle. For men, we estimate an average reduction of 3.8 disability-adjusted years (DALYs). The reduction in DALYs is relatively small for women. The gain in quality-adjusted life years (QALYs) is almost enough to offset all of the costs associated with program implementation for males and half of program costs for women.

Jorge Luis García
John E. Walker Department of Economics
Clemson University
205 Sirrine Hall
Clemson, SC 29634
US
jlgarci@clemson.edu

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and IZA
and also NBER
jjh@uchicago.edu

Related material are available at http://cehd.uchicago.edu/ece-health

1 Introduction

A substantial body of evidence documents that high-quality early childhood education boosts the skills of disadvantaged children.[1] Much of this research focuses on early childhood interventions with short-term followups. The few studies that analyze long-term outcomes primarily focus on labor market and criminal outcomes.[2] This paper examines the life-cycle health outcomes of an influential pair of essentially identical early childhood programs conducted in North Carolina that targeted disadvantaged children: the Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE), henceforth ABC/CARE. The programs were implemented using randomized trials. They begin early in participants' lives (at 8 weeks), and engage participants until age 5.[3]

Figure 1 summarizes the experimental treatment effects of ABC on the prevalence of risk factors for cardiovascular and metabolic diseases as measured by an epidemiological survey conducted when participants were in their mid-30s. There are pronounced differences in health treatment effects by gender. This paper projects estimated treatment effects at that age over the full life cycle using an adaptation of the Future Adult Model (FAM). Goldman et al. (2016) documents FAM. We examine the risk-reducing properties of high-quality early childhood education on six chronic conditions: cancer, lung disease, diabetes, heart disease,

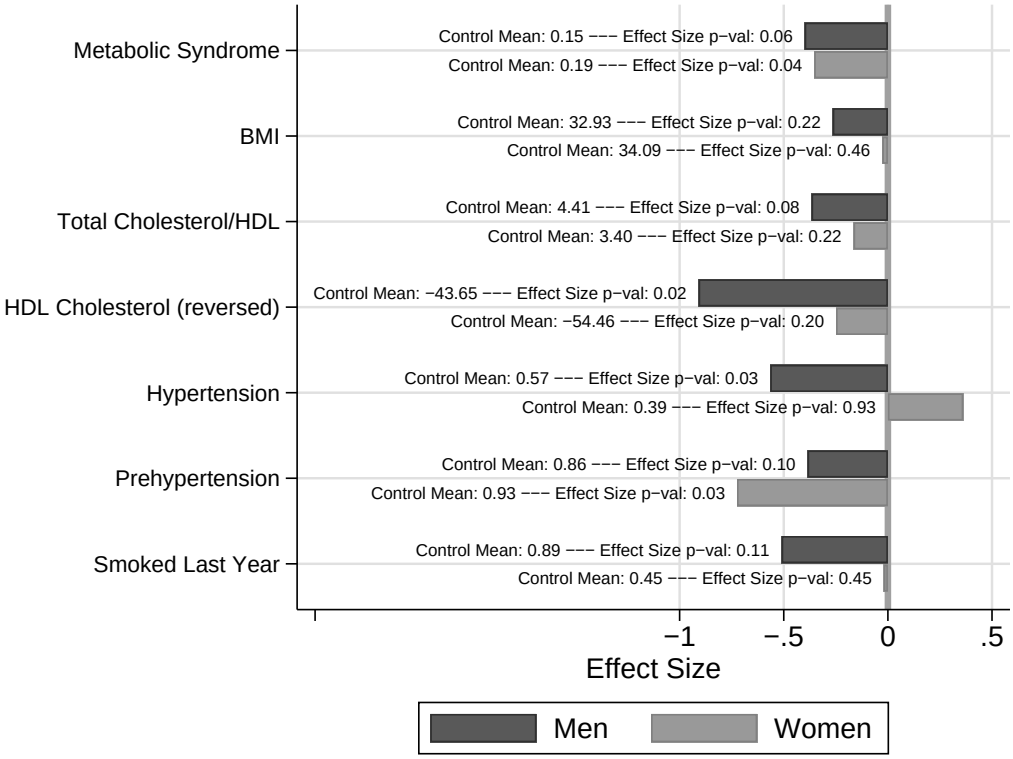
[1] See Cunha et al. (2006), Almond and Currie (2011), and Elango et al. (2016) for surveys.

[2] For example, Heckman et al. (2010), Havnes and Mogstad (2011), and Campbell et al. (2014).

[3] Both programs were launched in the 1970s. Their original goal was to boost cognition and promote schooling and attachment to society. Programs inspired by ABC/CARE have been (and are currently being) launched around the world. Sparling (2010) and Ramey et al. (2014) list numerous programs based on the ABC/CARE approach. The programs are: Infant Health and Development Program (IHDP) in eight different cities in the U.S. (Spiker et al., 1997); Early Head Start and Head Start. (Schneider and McDonald, 2007); John's Hopkins Cerebral Palsy Study in the U.S. (Sparling, 2010); Classroom Literacy Interventions and Outcomes (CLIO) study. (Sparling, 2010); Massachusetts Family Child Care Study (Collins et al., 2010); Healthy Child Manitoba Evaluation (Healthy Child Manitoba, 2015); Abecedarian Approach within an Innovative Implementation Framework (Jensen and Nielsen, 2016); and Building a Bridge into Preschool in Remote Northern Territory Communities in Australia (Scull et al., 2015). Current Educare programs in the U.S. are also based on ABC/CARE (Educare, 2014; Yazejian and Bryant, 2012). ABC had a second treatment stage from ages 5 to 8. Treatment was randomized using a second, independent randomization protocol. We do not analyze the second stage in this paper. The second stage had very weak effects on participants (see Campbell et al., 2013).

hypertension, and stroke. The program generates substantial lifetime benefits as measured by disability-adjusted life years, quality-adjusted life years, and mortality. García et al. (2020) use these estimates as one input into a cost-benefit analysis of the ABC/CARE program that also considers earnings, education, crime, and other outcomes. This paper examines the lifetime health benefits in much more detail than that paper and does not conduct a cost-benefit analysis.

Figure 1: ABC Health Effects at Mid-30s



Note: This figure describes the impact of ABC/CARE on mid-30s health. The effect size is calculated by dividing the treatment effect by the control-group standard deviation. The p-values are bootstrapped, nonparametric, and stratified by cohort, sex, and number of siblings at the baseline as in the original randomization protocol. We draw 1,000 bootstrap repetitions. Smoked Last Year: an indicator of having smoked tobacco during the year previous to the interview. Prehypertension: an indicator of systolic pressure > 120 or diastolic pressure > 80. Hypertension: an indicator of systolic pressure > 140 or diastolic pressure > 90. HDL Cholesterol (reversed): HDL or "good" cholesterol (mg/dl) multiplied by -1. Total Cholesterol/HDL: Total cholesterol to HDL cholesterol ratio. BMI: body-mass index. Metabolic Syndrome: an indicator of having metabolic syndrome.

FAM is a life cycle economic demographic microsimulation model for a nationally repre-

sentative sample: Panel Survey of Income Dynamics (PSID). We extract a subsample of ABC/CARE eligibles. We combine the FAM subsample with the ABC/CARE experimental sample and use the methodology of García et al. (2020) to project lifetime health benefits. [4]

We find substantial life-cycle effects on cancer, heart disease, stroke, and mortality for both males and females. Across the life-cycle, treatment group females are less likely to have a stroke than their treatment group counterparts. By age 70, treatment group males are half as likely to have a stroke than control group counterparts. Similar results hold for heart disease, lung disease, and cancer. By age 50, a significant difference between the likelihood of dying due to any of these diseases emerges between treatment and control group participants, both male and female. For males after age 60, this difference grows substantially. At age 70, control group males are almost four times as likely to die as their treatment group counterparts. Results for women are less stark. For males, we calculate a statistically significant average reduction of 3.8 disability-adjusted years, while for females the gain is small and imprecise. For males, the increase in quality-adjusted life years more than offsets all program costs (i.e., if the only benefit of the program had been the improvement in QALYs, the program would have almost paid for itself). For females, the increase in quality-adjusted life years offsets nearly half of the program costs.

The paper proceeds as follows. Section 2 describes the ABC/CARE program and the data collected on its participants. Section 3 discusses our adaption of the FAM model to forecast life-cycle chronic diseases for ABC/CARE participants. Section 4 reports and discusses our findings. Section 5 concludes.

[4] García et al. (2020), show that, under the null hypothesis of input exogeneity, it is possible to formulate additional tests for this methodology. It is also possible to test for input exogeneity. Unfortunately, we can only perform those tests for outcomes for which there are two observation periods during adulthood, which is not the case of health. García et al. (2020) test for and do not reject exogeneity using labor income as an outcome. Given the prior tests, we assume that, conditioning on regressors, inputs are exogenous.

2 ABC/CARE Program and Data Description

ABC/CARE is a randomized early childhood intervention with longitudinal follow-up data from birth until adulthood. Data on cognitive skills, socio-emotional skills, family environment, and self-reported health of control and treatment group children were collected annually during the duration of the program and then periodically until participants reached their mid-30s. ABC/CARE is a widely emulated early childhood program. Despite larger sample sizes and multi-site designs in other randomized programs, few have compiled longitudinal data on health outcomes until adulthood (Elango et al., 2016). Long-term outcomes are especially important for health since many chronic conditions manifest later in life after years of sustained behavior.

The goal of the Carolina Abecedarian Project (ABC) and Carolina Approach to Responsive Education (CARE), ABC/CARE, was to prepare children for school socially, cognitively, and academically by promoting language and cognitive development through center-based care. There was no focus on adult health, yet we find substantial health impacts. The interactive curriculum provided an educational environment with small student-staff ratios and small-group learning. The program also provided nutritious meals and medical checkups for participants.

ABC and CARE recruited four and two cohorts, respectively, of disadvantaged children born in Chapel Hill, North Carolina between 1972 and 1980. Potential participants were referred to researchers by local social service agencies and hospitals at the beginning of the mother's last trimester of pregnancy. Eligibility was based on a High Risk Index[5] developed by the Frank Porter Graham Center (FPG) at the University of North Carolina at Chapel Hill.

[5]The index weighted the following variables: maternal and paternal education, family income, father's presence at home, lack of maternal relatives in the area, siblings behind appropriate grade in school, family on welfare, father in unstable job, low maternal IQ, low siblings' IQ, social agency indicates that the family is disadvantaged, one or more family members has sought a form of professional help in the last three years, and any other special circumstance detected by program staff.

Eligible mothers were 20 years old on average, 74% of fathers were absent, and 94% of the sample was African-American.

The final ABC sample consisted of 114 subjects. 58 were in the treatment group. 56 were in the control group. CARE consisted of 65 families, with 25 in a family education treatment group, 23 in the control group, and 17 in a center-based childcare treatment group. Six and five subjects withdrew from the ABC and CARE samples, respectively.

Children were randomized into treatment and control groups using child pairs matched based on family background. All subjects received diapers and formula for the first six months, and treatment group subjects received additional daily health screenings. From the ages of 0 to 5, treatment group subjects received cognitive and social stimulation for eight hours a day in center-based care.[6] Even though CARE subjects also received home visits from the ages of 0 to 5, this component was shown to have very weak estimated effects using a second treatment arm that received home visits only (Campbell et al., 2013). Previous analyses justify merging the treatment groups of ABC and the main treatment arm of CARE.[7]

Follow-up data collection of ABC/CARE subjects occurred at ages 12, 15, 21, and 30. Various education, employment, health, crime, and family structure measures were collected through both administrative and self-reported channels. Additionally, data from a full-medical sweep of participants in their mid-30s is also available. Our analysis leverages this data. We exclude five outlier subjects with mid-30s BMI greater than 50 in the final analysis,

[6] ABC/CARE also included school-age intervention from the ages of 5 to 8 focused on reading and math, which was found to have no effect (Campbell et al., 2013).

[7] We acknowledge that dropping one of the treatment arms is a possible concern, as pointed out by Muralidharan et al. (2019), although people were randomly assigned to treatment arms. We are faced with a trade-off. We could analyze the ABC sample only as previous studies have done and not recognize that CARE was a continuation of the ABC/CARE program or we could include CARE. When including CARE, however, it is not sensible to include the treatment arm that only received home visits. Home visits were a very minor element of the program and comparisons to the control group would be imprecise given all of the estimation steps in our procedure and the small-sample size of the home-visits treatment arm. A fully-interacted model would be imprecisely estimated in our complex setting. Not analyzing the school-stage randomization is not a concern. School-stage treatment was assigned through a completely new randomization so critiques like those in Muralidharan et al. (2019) do not apply.

leaving a final sample of 83 individuals. Outliers that are this extreme are problematic because they generate imprecise predicted treatment effects because of our small sample size used to initialize the forecasts.[8]

3 Forecasting Chronic Diseases for ABC/CARE

We first formally state the Future Adult Model (FAM), the model that we use for predicting life-cycle treatment effects. We then apply it.

3.1 Notation and Model Development

Let $\mathcal{M}$ denote a set of possible health states, some of which can be absorbing. Let $\mathcal{A} := [0, \ldots, \bar{A}]$ index ages, where $\bar{A}$ is the last age for which we construct forecasts. We define $h_{a,m,m'}$ as the probability of transitioning from state m to state m' at age $a \in \mathcal{A}$, where $m, m' \in \mathcal{M}$. We drop individual subscripts to avoid notational clutter.

We denote a transition from m to m' at age a by $D_{a,m,m'} = 1$. If this transition does not occur, $D_{a,m,m'} = 0$. We let $\tilde{D}_{a,m}$ be the indicator of occupancy of state m at age a. $\tilde{D}_{a,m}$ is a generic entry in the the vector of age-a state occupancy indicators denoted by $\tilde{\boldsymbol{D}}_a$. $\tilde{\boldsymbol{D}}_0$ denotes the vector of initial state-occupancy conditions.

The probability of occupying state $m \in \mathcal{M}$ at age $a \in A$ is assumed to be generated by an index threshold-crossing model:

$$I_{a,m} = \boldsymbol{1}\left(\tilde{\boldsymbol{D}}_0 \geq 0\right)\boldsymbol{\Omega}_m + \boldsymbol{1}\left(\tilde{\boldsymbol{D}}_{a-1} \geq 0\right)\boldsymbol{\Lambda}_m + \boldsymbol{W_a}\boldsymbol{\beta}_m + \boldsymbol{B}\boldsymbol{\alpha}_m + \tau_{a,m} + \varepsilon_{a,m}, \quad (1)$$

[8]Note that we only drop them in the forecast exercises. Individuals with BMI over fifty are still included in Figure 1. Despite the inclusion of five relatively unhealthy individuals, as measured by their BMI, in the treatment group (four females and one male; two in the ABC sample and three in the CARE sample), the mid-30s treatment effects on health outcomes are economically and statistically significant.

where $\tilde{D}_{a,m} = \boldsymbol{1}\left(I_{a,m} \geq 0\right)$, $\boldsymbol{1}\left(\tilde{\boldsymbol{D}}_0\right)$ is a vector of indicators of mutually exclusive initial conditions with associated coefficients $\boldsymbol{\Omega}_m$, $\boldsymbol{1}\left(\tilde{\boldsymbol{D}}_{a-1} \geq 0\right)$ is a vector of mutually exclusive previous-period outcomes associated coefficients $\boldsymbol{\Lambda}_m$. Both $\tilde{\boldsymbol{D}}_0$ and $\tilde{\boldsymbol{D}}_{a-1}$ can be affected by treatment. $\boldsymbol{W}_a$ denotes contemporaneous variables that can be affected by treatment with associated coefficients $\boldsymbol{\beta}_m$. $\boldsymbol{B}$ is the vector of eligibility conditions with associated coefficients $\boldsymbol{\alpha}_m$. $\tau_{a,m}$ is age.[9] $\varepsilon_{a,m}$ is a serially uncorrelated shock which is assumed to be uncorrelated with all of the right-hand side observables. In practice, when analyzing discrete outcomes, we assume that it is a unit normal random variable.

Instead of modeling initial conditions, we directly condition on them in Equation (1). The assumption of uncorrelatedess of $\varepsilon_{a,m}$ across ages, health states, and subjects together with normality allows us to conduct separate estimation of the coefficients characterizing Equation (1) for each $m \in \mathcal{M}$ using maximum likelihood. We estimate $[\boldsymbol{\Omega}_m, \boldsymbol{\Lambda}_m, \boldsymbol{\beta}_m, \boldsymbol{\alpha}_m, \tau_{a,m}]$ for each $m \in \mathcal{M}$.

The probability of occupying various discrete states is generated in FAM using Equation (1). When simulating the model to forecast the health outcomes for the ABC/CARE subjects, we use their observed age-30 conditions as initial conditions $\tilde{\boldsymbol{D}}_0$. Most of these initial conditions relate to health, but we also include initial conditions related to economic status.

This model easily accommodates absorbing states. Thus, $m \in \mathcal{M}$ is an absorbing state if $\tilde{D}_{a,m} = 1$ implies that $\tilde{D}_{a',m} = 1\ \forall a' \in \mathcal{A}$ with $a' \geq a$. Once an individual reaches an absorbing stage, their model in Equation (1) is no longer used to generate transitions. Death is clearly an absorbing state.

Ordered and unordered outcomes are also easily accommodated using standard methods in discrete choice. An example of the former is the level of psychological distress (e.g., low, medium, high). An example of the latter is labor force status (e.g., labor force status is

[9] In our empirical analysis, we approximate the coefficients on $\tau_{a,m}$ using splines with knots at ages 35, 45, 55, 65, and 75.

categorized as out of labor force, unemployed, working part time, or working full time). We also model continuous outcomes so that $I_{a,m}$ is observed and the variables in $\tilde{\boldsymbol{D}}_a$ are replaced by observed counterparts. An example of a continuous outcome is body-mass index (BMI). We consider other inputs that are related to health but are not health outcomes *per se*. These include labor force participation, relationship status, and childbearing models. They are estimated within FAM to account for the socioeconomic disadvantage of participants. We take the labor income forecast of García et al. (2020) as an input for the heath predictions.

Tables 1 to 3 list the variables determining each of the states and health and economic outcomes that we analyze. For each of the outcomes we list: (1) the outcome itself; (2) the variable type—e.g., absorbing state, binary outcome, continuous outcome; (3) initial health state occupancies and other outcomes— $\tilde{\boldsymbol{D}}_0$ in Equation (1); (4) lagged health-state occupancies—$\tilde{\boldsymbol{D}}_{a-1}$ in Equation (1); (5) and (6) other health and economic outcomes used to determine the outcome of interest—$\boldsymbol{W}_a$ in Equation (1); and (7) background variables—$\boldsymbol{B}$ in Equation (1).

FAM belongs to one of the two general classes of forecasting models employed in health economics: state-based models and potential impact fractions models.[10] In particular, FAM operationalizes state-dependency through a first-order Markov stochastic structure as in Briggs et al. (2006). More complete models allow for complete history dependency like time elapsed after first diagnosis. However, this history is scarce and difficult to include in scenarios like ours, which is a general problem in the literature (Richardson et al., 2011). Goldman et al. (2016) justify the modeling choices by appealing to research and the advice of clinicians and other medical professionals. They also document that the FAM performs well in fitting full population means of the forecasted outcomes. The first-order Markov stochastic nature of the model is a main assumption and we test it below.

[10] See van Baal and Boshuizen (2019) for a review of health forecasting models.

Table 1: Determinants of Equation (1) for Different Outcomes

(1) Outcome	(2) Variable Type	(3) $\tilde{\boldsymbol{D}}_0$ Initial Conditions	(4) $\tilde{\boldsymbol{D}}_{a-1}$ Past Outcomes	(5) $\boldsymbol{W}_a$ Health Behaviors	(6) $\boldsymbol{W}_a$ Economic Outcomes	(7) $\boldsymbol{B}$ Demographics
Heart Disease	Absorbing	Childhood Economic Environment Education Asthma	Hypertension Diabetes	Smoking BMI Physical Activity		Race Ethnicity Age Gender
Hypertension	Absorbing	Childhood Economic Environment Education	Diabetes	Smoking BMI Physical Activity		Race Ethnicity Age Gender
Stroke	Absorbing	Childhood Economic Environment Education	Heart Disease Hypertension Diabetes Cancer	Smoking BMI Physical Activity		Race Ethnicity Age Gender
Lung Disease	Absorbing	Childhood Economic Environment Education Asthma		Smoking BMI Physical Activity		Race Ethnicity Age Gender
Diabetes	Absorbing	Childhood Economic Environment Education		Smoking BMI Physical Activity		Race Ethnicity Age Gender
Cancer	Absorbing	Childhood Economic Environment Education		Smoking BMI Physical Activity		Race Ethnicity Age Gender
Mortality	Absorbing	Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer Functional Status	Smoking Binge Drinking		Race Ethnicity Age Gender
Functional Status	Ordered	Childhood Economic Environment Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Physical Activity Functional Status		Race Ethnicity Age Gender
Smoking	Binary	Childhood Economic Environment Education	Heart Disease Lung Disease Diabetes	BMI Binge Drinking Physical Activity Psychological Distress		Race Ethnicity Age Gender

Note: This table provides details on the empirical specification of Equation (1) for the different outcomes that we consider.

Table 2: Determinants of Equation (1) for Different Outcomes, Continued

(1) Outcome	(2) Variable Type	(3) $\tilde{\boldsymbol{D}}_0$ Initial Conditions	(4) $\tilde{\boldsymbol{D}}_{a-1}$ Past Outcomes	(5) $\boldsymbol{W}_a$ Health Behaviors	(6) $\boldsymbol{W}_a$ Economic Outcomes	(7) $\boldsymbol{B}$ Demographics
BMI	Continuous	Childhood Economic Environment Education		BMI	Marital Status	Race Ethnicity Age Gender
Binge Drinking	Binary	Childhood Economic Environment Education	Marital Status	Binge Drinking		Race Ethnicity Age Gender
Physical Activity	Binary	Childhood Economic Environment Education	Marital Status	Physical Activity		Race Ethnicity Age Gender
Psychological Distress	Ordered	Childhood Economic Environment Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Physical Activity Psychological Distress Functional Status		Race Ethnicity Age Gender
Childbearing	Ordered	Mother's Education Education	Cancer	Marital Status	Labor Force Participation Number of Children	Race Ethnicity Age Gender
Paternity	Ordered	Mother's Education Education			Labor Force Participation Marital Status Number of Children	Race Ethnicity Age
Marital Status	Binary	Mother's Education Education			Labor Force Participation Earnings Marital Status Number of Children	Race Ethnicity Age Gender
Partner Mortality	Binary	Education				Race Ethnicity Age Gender

Note: This table provides details on the empirical specification of Equation (1) for the different outcomes that we consider.

Table 3: Determinants of Equation (1) for Different Outcomes, Continued

(1) Outcome	(2) Variable Type	(3) $\tilde{\boldsymbol{D}}_0$ Initial Conditions	(4) $\tilde{\boldsymbol{D}}_{a-1}$ Past Outcomes	(5) $\boldsymbol{W}_a$ Health Behaviors	(6) $\boldsymbol{W}_a$ Economic Outcomes	(7) $\boldsymbol{B}$ Demographics
Labor Force Participation	Unordered Categorical	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Supplemental Security Income Claiming Earnings Marital Status	Race Ethnicity Age Gender
Full-time Employment	Binary	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Supplemental Security Income Claiming Earnings	Race Ethnicity Age Gender Marital Status
Disability Insurance Claiming	Binary	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Earnings	Race Ethnicity Age Gender
Social Security Claiming	Absorbing	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Earnings Marital Status	Race Ethnicity Age Gender
Supplemental Security Income Claiming	Binary	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Supplemental Security Income Claiming Earnings Marital Status	Race Ethnicity Age Gender
Health Insurance Type	Unordered Categorical	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Earnings Marital Status Health Insurance Type	Race Ethnicity Age Gender
Nursing Home Residency	Binary	Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Functional Status	Nursing Home Residency Widowhood	Race Ethnicity Gender Age

Note: This table provides details on the empirical specification of Equation (1) for the different outcomes that we consider. Functional status describes the ability of an individual to perform normal activities to meet basic needs, fulfill usual roles in a job, and maintain normal health.

3.2 Simulating the Model

We estimate the parameters of the models in Equation (1) on a sample eligible for ABC/CARE. With these estimates, we predict a future health trajectory for each ABC/CARE subject by initializing it with the individual initial conditions. To match the biennial structure of the Panel Study of Income Dynamics (PSID) data used to estimate the models, the simulation for predicting proceeds in two-year increments.[11]

Among the ABC/CARE subjects simulated in FAM, the years of completion of the age-30 interview range from 2003 to 2009. FAM's two-year time step only allows the simulation of even or odd years. For this reason, we run the simulation twice—once for the ABC/CARE subjects entering in odd years and again for the ABC/CARE subjects entering in even years.

The simulation model uses assumptions regarding the normal retirement age, future improvements in mortality, and real medical cost growth as inputs. The normal retirement age is assumed to be 67 for all ABC/CARE subjects.

The FAM mortality model represents mortality rates in 2009. The estimated mortality probabilities are reduced in simulated future years to represent improvements in mortality from sources such as medical innovation that are not included in the model. There are different adjustment factors for populations under and over the age of 65. The mortality reduction factors are taken from the intermediate cost mortality projections in the 2013 Social Security Trustee's Report.

[11]The end of each two-year step is designed to occur on July 1st to allow for easier matching with population forecasts from the Social Security Administration (SSA).

3.3 Data Sources for Estimation and Simulation

FAM uses data from ABC/CARE surveys to set the initial conditions of the simulation model. The state-occupancy model parameters are estimated using the 1997 to 2013 waves of the PSID. We use the PSID because it is the longitudinal sample with the longest surveyed life cycle measures. It has extensive information concerning demographics, economic outcomes, health care access, health outcomes, and health behaviors (such as smoking history, alcohol consumption, and exercise habits). We restrict the PSID to heads of households age 25 and older because these subjects respond to the most comprehensive questions. The FAM forecasts of BMI are based on self-reported height and weight.[12] In ABC/CARE, BMI is measured as part of the mid-30s medical examination. Additionally, BMI is self-reported in the PSID at age 30, while it is measured during the mid-30s for the ABC/CARE sample. We supplement the PSID with the the National Health and Nutrition Examination Survey (NHANES), which has longitudinal measures and self reports of BMI, to predict BMI at age 30 for the ABC/CARE individuals. Since the PSID does not follow individuals in nursing homes or other long-term care facilities, we supplement the PSID with the HRS when estimating mortality models. For the HRS, we use all cohorts in the dataset created by RAND, version O. The NHANES includes both self-reported height and weight as well as epidemiological measures of BMI, which is why it is employed for interpolating physical measures from self-reported measures.

[12] That is how it is reported in the PSID.

Table 4: Summary of Data Sources

	ABC/CARE	PSID	HRS	NHANES
Ages used	0-34	25+	50+	30-40
Years used	- -	1997-2013	1998-2012	2002-2010
Longitudinal	✓	✓	✓	
Time Intervals of Data Collection	Ages 0-8, 12, 15, 21, 30, 34	Biennial	Biennial	Annual
Demographic Outcomes	✓	✓		✓
Economic Outcomes	✓	✓	✓	
Health Outcomes	✓	✓	✓	✓
Health Behaviors	✓	✓	✓	
Health Expenditures	✓		✓	
Family Outcomes	✓	✓		
Includes Institutionalized Individuals		✓		
Models	Initializing all models	Health	Mortality, widowhood, nursing home residency	BMI

Note: This table compares the main features of the auxiliary datasets used in simulating life-cycle health outcomes of ABC/CARE subjects. We restrict the PSID to heads of households aged 25 and older because these subjects respond to the largest set of questions.

3.3.1 Variable Construction and Imputations

Some of the initializing variables are not available for all ABC/CARE subjects at the required ages for FAM and are imputed using the data sources in Table 4. The imputations made are described in Table 5.

Table 5: Imputation of Model Inputs

Input	Subjects with Missing Data	Models Requiring Input	Variables Used to Impute	Method Used to Impute
Mother's Education Level	CARE subjects	Marital status and childbearing	Race, ethnicity, education, disease conditions, employment status, presence of a health-related work limitation, and a self-report of whether or not the subject was "poor" as a child	Ordered probit model is constructed using PSID data of subjects age 30 and 31 born between 1945 and 1981.
Socioeconomic Status of Parents	All subjects	Numerous models		Assume all subjects were "poor" given program eligibility.
Race	All subjects	All models		Assume no participants are Hispanic or Latino.[1]
Smoking and Employment Status	1 subject	Health states, marital status, childbearing, DI and SSI benefits, health insurance category		Multinomial logit model to estimate joint probability of each comibnation of smoking and employment amongst unemployed PSID subjects age 25 to 35.
Binge Drinking	1 subject	Mortality, smoking		Binary probit model using PSID data of subjects age 25 to 35.
BMI	All subjects	Health states, functional status, employment, and smoking	Self-reported height and weight at age 30 (CARE) or age 34 (BMI)	Covariate values from PSID age 30-34 data in 2002-2013 are used to impute measured BMI values for PSID respondents using estimated models based on NHANES and the method in Courtemanche et al. (2015). PSID is then used to estimate a model mapping imputed measured BMI at ages 33-40 to selfreported BMI at ages 30-32. This imputation is applied to ABC/CARE subjects with a health interview at least one year after their age 30 interview.
Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs)	All subjects	Benefits claiming, mortality, employment status, insurance category, and nursing home residency	If the subject has a physical or nervous condition that keeps them from working	Ordered Probit model estimated on PSID respondents aged 25 to 35.
DI and SSI benefits	All subjects	Employment status, insurance category, and Medicare enrollment	Single question asking about receipt of any kind of benefits	Multinomial logit model to estimate the joint probability of each combination of DI and SSI claiming using PSID respondents age 25 to 35 on benefits.

[1]Census data on Hispanics in North Carolina was not available for 1970 and 1980, but Hispanic migration into this state is more recent than in other regions, and as late as 1990, only 2% of the North Carolina poor were Hispanic (Johnson, 2003).

Note: This table summarizes the models estimated in order to impute necessary outcomes in the FAM Model. ADLs include walking, dressing, eating, bathing or showering, getting in and out of bed or a chair, and using the toilet, including getting to the toilet. IADLS include preparing meals; shopping for toiletries and medicine; managing money; using the phone; doing heavy housework; and doing light housework or housecleaning.

3.4 Specification Tests for the First-order Markov Assumption

The FAM model assumes that a vector first-order Markov process governs transitions with a variable state space depending on the analyzed outcome. In Equation (1), only $\tilde{\boldsymbol{D}}_{a-1}$ and the initial state value enter the model. For heart disease, hypertension, and stroke, we compare the fit of Equation (1) with the following second-order model:

$$\begin{aligned} I_{a,m} &= \boldsymbol{1}\left(\tilde{\boldsymbol{D}}_0 \geq 0\right)\boldsymbol{\Omega}_m + \boldsymbol{1}\left(\tilde{\boldsymbol{D}}_{a-1} \geq 0\right)\boldsymbol{\Lambda}_m + \left(\tilde{\boldsymbol{D}}_{a-2} \geq 0\right)\tilde{\boldsymbol{\Lambda}}_m \\ &+ \boldsymbol{W_a}\boldsymbol{\beta}_m + \boldsymbol{B}\boldsymbol{\alpha}_m + \tau_{a,m} + \varepsilon_{a,m}. \end{aligned} \tag{2}$$

We use a likelihood ratio test to test the null hypothesis $H_0 : \tilde{\boldsymbol{\Lambda}}_m = \boldsymbol{0}$ in Equation (2). Table 6 show the results from these tests conducted on the samples used to make the forecasts reported in this paper. For the health states analyzed, we do not reject the null hypothesis. We do not perform these tests for lung disease, cancer, and diabetes because sample limitations do not give us $\tilde{\boldsymbol{D}}_{a-2}$ for these diseases.

Table 6: Tests Comparing First-Order and Second Markov Processes for Disease State-Occupancy Specifications

Disease	LR Statistic	Degrees of Freedom	p-value
Heart Disease	2.18	2	0.71
Hypertension	0.05	1	0.83
Stroke	3.94	4	0.14

Note: This table presents likelihood ratios contrasting the models in Equations (1) and (2) by testing the null hypothesis $H_0 : \tilde{\boldsymbol{\Lambda}}_{a,m} = 0$. The variables included in the right-hand-side of each model are in Tables 1 to 3.

3.5 Inference

We study the life-cycle trajectories after age 30 for six chronic diseases and associated treatment effects. This produces hundreds of age-wise treatment effects across diseases. Summarizing these effects in an interpretable way is challenging. Following García et al. (2018) we construct combining functions that, within disease categories, count the proportion of treatment effects that have the same sign. We generate standard errors for these counts.

Formally, consider a block of outcomes $\mathcal{J}_\ell, \ell \in \{1, \ldots, L\}$, with cardinality C_ℓ and associated treatment effects $\Delta_1, \ldots, \Delta_{C_\ell}$. For the case of diabetes, a block is a set of indicators of prevalence of diabetes between ages 30 and 40, 30 and 50, 30 and 60, and 30 and 70. For each of these blocks, we observe the prevalence of diabetes for each age. We construct similar blocks with the rest of the chronic diseases.

Treatment effects, Δ_j, can be either beneficial or detrimental. The interpretation placed on the sign of the combining function is evident from the context. The count of positive-valued treatment effects within block $\mathcal{J}_\ell$ is

$$D_\ell = \sum_{j=1}^{C_\ell} \mathbf{1}(\Delta_j > 0). \tag{3}$$

We use the proportion of outcomes with $\Delta_\ell > 0$ as our combining function: D_ℓ/C_ℓ. Under the null hypothesis of no treatment effect for the block of outcomes indexed by $\mathcal{J}_\ell$, and assuming the validity of asymptotic approximations, the mean of D_ℓ/C_ℓ is centered at $\frac{1}{2}$.[13] We compute the fraction D_ℓ/C_ℓ and the corresponding bootstrapped empirical distribution to obtain a p-value. The bootstrap procedure accounts for dependence in unobservables across outcomes (within blocks) in a general way.

[13] Campbell et al. (2014) establish the validity of asymptotic approximations for the ABC sample.

4 Empirical Results

We plot the probabilities of incidence of mortality and disease by 5-year bins from age 30 to 75 over the life-cycle for cancer, lung disease, diabetes, heart disease, hypertension, and stroke. Disease incidence is defined as a diagnosis or death in order to account for disease-free survival. We perform inference—at ages 40, 50, 60, and 70—on combining functions for the various chronic diseases. The combining functions are the proportion of years starting at age 30 which exhibit a positive treatment effect for each chronic disease. For example, at age 40, the combining function for lung disease is the proportion of years with a positive treatment effect on this chronic disease between ages 30 and 40. We calculate one-sided p-values using 1,000 bootstrap samples of the FAM under the null hypothesis that D_ℓ/C_ℓ is equal to 50%. If one were to plot the trajectories for all 1,000 bootstrap samples, the p-value is the percentage of simulations where the majority of the control group trajectory is above that of the treatment group. This choice of inference emphasizes the persistence of treatment effects over the life-cycle as opposed to treatment effect magnitudes at a single age. It detects a consistent life cycle pattern of beneficial outcomes. Life-cycle trajectories are plotted by 5-year bins while inference is performed using blocks of annual data. This can lead to slight discrepancies between the actual D_ℓ/C_ℓ and expected D_ℓ/C_ℓ based on the smoothed plot. We also display bootstrapped standard errors for the 5-year bins.

The combining functions are statistically significant at standard levels for heart disease, stroke, cancer, and mortality over most of the lifecycle for both genders. Females also have statistically significant combining functions for diabetes up to age 50. For males, diabetes is the only disease with adversely significant combining functions throughout the life-cycle. An explanation for this is that prevalence of familial history suggests higher genetic predisposition for diabetes for the male treatment group. This is the only family history variable for which there is a difference between the treatment and control groups.

Once we condition on family diabetes history, the adverse effect disappears. Treatment improves the epidemiological assessments in Figure 1 that we use as the starting point for our forecasts. Thus, the only reason the adverse effect of treatment on diabetes arises in the longitudinal profiles is the initial difference in diabetes which continues through most of the life cycle in the simulation.[14] We discuss this further below. The program has a positive effect on nondiabetic cardiovascular conditions that are mediated by the health improvements at mid-30s, as shown by the epidemiological assessments in Figure 1.

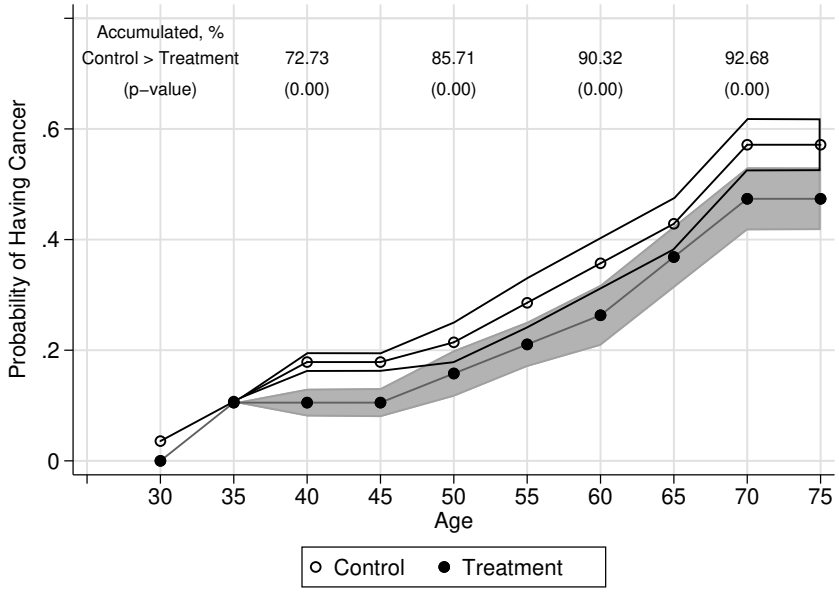
We now describe the life-cycle trajectory for each chronic condition. There are statistically significant empirical combining functions for males up to age 70 and over the entire life-cycle for females in Figure 2. Female control and treatment incidence rates converge at age 35, after which treatment group incidence rates remain slightly lower than those of the control group. Near-zero p-values at all ages indicate a similar trend across all bootstraps. Treatment-group males exhibit a lower incidence rate than control-group males until age 58. There is a second crossover just after age 65, and both groups converge at age 75. Since age is one of the biggest risk factors for cancer, it is unsurprising that the combining function becomes insignificant at the end of the life-cycle.

There are no statistically significant combining functions for lung disease. See Figure 3. However, life-cycle patterns differ noticeably by gender. Female treatment and control trajectories closely mirror one another. After age 35, incidence rates for treatment-group males are lower than those of control-group males by 15% to 20% throughout the life-cycle. Despite the statistically insignificant combining functions for males, the trajectories in Figure 3 show that the profiles for both the treatment and control groups are precisely estimated. A main input (but not the sole input) that enters into our forecast of this chronic disease is smoking behavior. In the control group, 89% of participants report they smoked during the last year

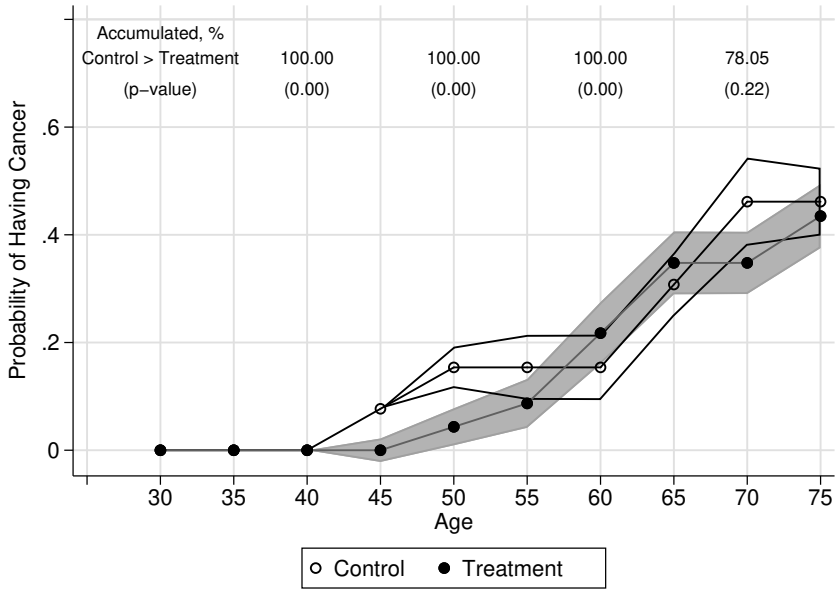
[14]Health experts state diabetes and other heart conditions such as heart disease, stroke, and hypertension share common risk factors. However, few genetic risk factors that co-modulate both conditions have been identified thus far (Sousa et al., 2011).

Figure 2: Life-cycle Trajectories of Cancer by Gender and Treatment Group with Associated Point-wise Confidence Intervals

(a) Female



(b) Male



Note: The figures plot the proportion of treatment and control subjects who either have the disease or die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual contracts the disease or dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect as well as standard errors. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation, which we also use to calculate standard errors. Combining functions are highly significant across the life-cycle for both genders.

during the mid-30s interview. In the treatment group, the analogous figure is 72%. The treatment-control difference is statistically significant at the 10% level (see Figure 1).

Significantly more treatment-group males have a maternal history of diabetes.[15] Scott et al. (2007) and Zeggini et al. (2008) identify a heritability component to diabetes in addition to environmental and behavioral factors.[16] This suggests that males assigned to the treatment group have a higher genetic predisposition to diabetes. Therefore, our estimation of the life-cycle trajectories of diabetes accounts for familiar diabetes history, by conditioning on mother, father, and siblings diabetes history.

There are statistically significant combining functions for females up to age 50 and perverse and significant combining functions for males over the lifecycle in Figure 4. Treatment-group females have lower incidence of diabetes from age 30 to 45, after which the female treatment trajectory remains marginally higher than those of the control. Combining functions that are statistically significant at ages 40 and 50 are insignificant at ages 60 and 70. Thus, there are consistent early life-cycle treatment effects for females bootstrap samples. Beneficial effects eventually emerge for males because the program improves several other health conditions, despite the diabetes predisposition in the treatment group.

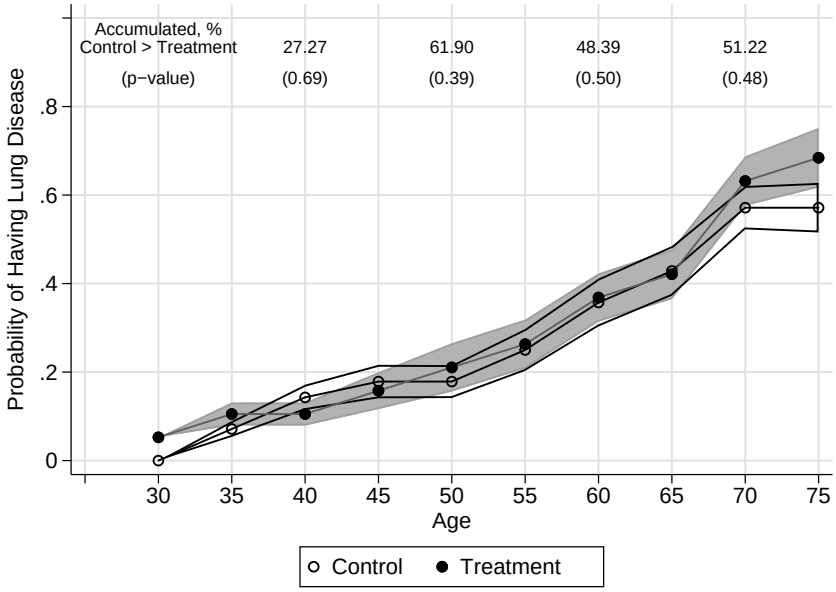
The combining functions for heart disease are highly significant up to age 60 for females and over the entire life-cycle for males (Figure 5). Treatment-group females have consistently lower incidence rates until age 50, after which their incidence rates are slightly higher than that of control. Combining functions are statistically significant until age 60. Treatment-group males have lower incidence rates than the control group over the life-cycle, and by a wider margin than females. Though male incidence rates converge at age 65, treatment group

[15]This can occur by chance even in a randomized control trial. 26% and 15% of treatment and control males have a maternal history of diabetes, respectively. There is also one treatment male with a history of paternal diabetes and one treatment male with a diabetic sibling.

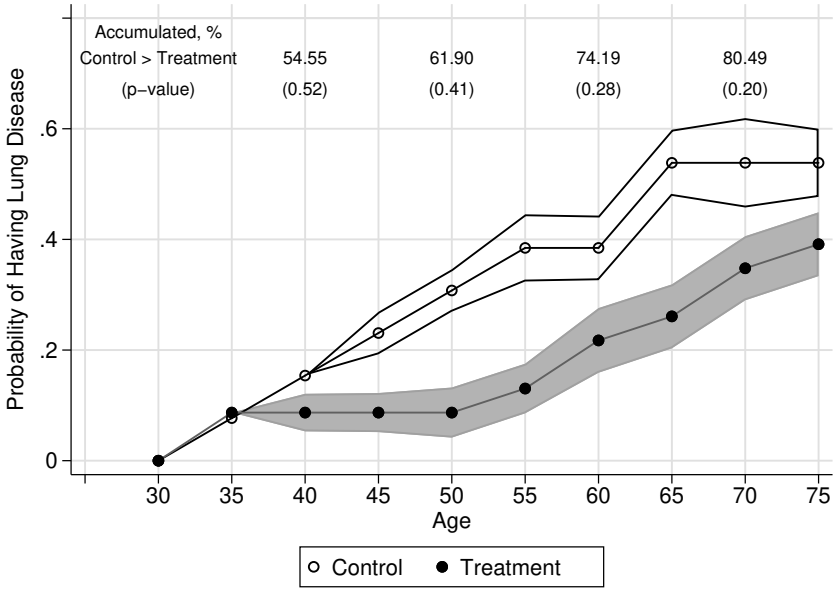
[16]Treatment subjects also have a high proportion of PMP-based diabetes by their midlife survey—17% of treatment males and no control males. Sabatier et al. (2002) find that patients with type 1 diabetes, which is genetically inherited, have elevated levels of PMP compared to type 2 diabetics.

Figure 3: Life-cycle Trajectories of Lung Disease by Gender and Treatment Group

(a) Female



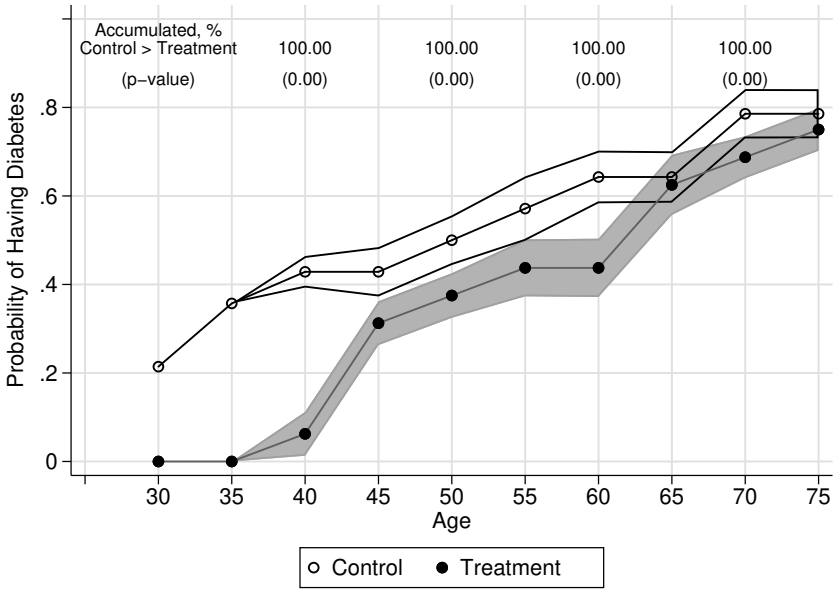
(b) Male



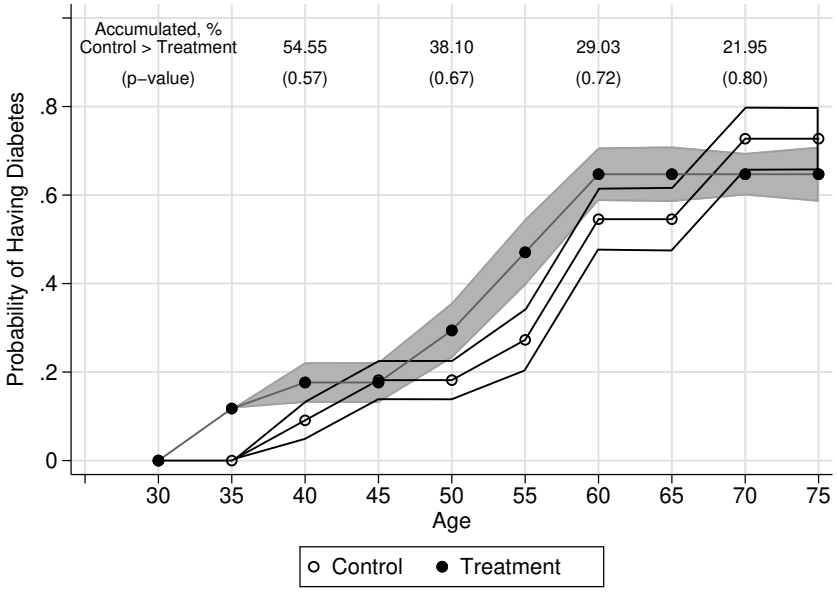
Note: The figures plot the proportion of treatment and control subjects who either have the disease or die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual contracts the disease or dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect as well as standard errors. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation, which we also use to calculate standard errors. There are no significant combining functions for lung disease.

Figure 4: Life-cycle Trajectories of Diabetes by Gender and Treatment Group with Associated Point-wise Confidence Intervals

(a) Female



(b) Male



Note: The figures plot the proportion of treatment and control subjects who either have the disease or die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual contracts the disease or dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect as well as standard errors. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation, which we also use to calculate standard errors. There are highly significant combining functions up to age 50 for females. The initial conditions for forecasting account for familiar diabetes history, by conditioning on mother, father, and siblings diabetes history

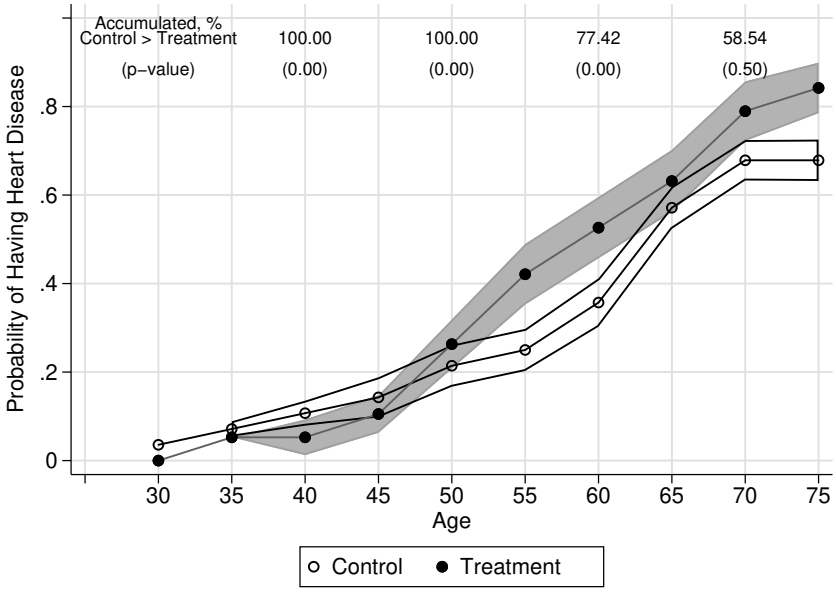
rates plateau at 50% while control group rates continue to increase. Combining functions equal 100% for all ages and have high significance, indicating consistent trajectories across bootstraps.

There are no statistically significant treatment effects for hypertension. The entire sample is diagnosed by age 65 in Figure 6, which is likely a predisposition of the individuals in the sample to contract this disease which the treatment fails to prevent. Treatment-group females have slightly higher incidence rates until age 50, after which control-group females have much higher incidence rates until age 65. There are perversely significant combining functions at ages 40 and 50. They become insignificant after age 60. There is a similar trend for males: treatment-group males have marginally higher incidence rates until age 40, after which control-group males have much higher incidence rates before convergence at age 65. There is an adversely significant combining function at age 40, but the estimates are statistically insignificant for the remainder of the life-cycle. Therefore, the difference in the incidence of hypertension is insignificant in the long-run. This can be attributed to the fact that almost the entire sample contracted hypertension by age 65. This is unsurprising given that African-Americans have the highest rates of hypertension globally, with estimated rates of 45.0% and 46.3% for adult African-American males and females, respectively, in 2011 to 2014 (Benjamin et al., 2017).

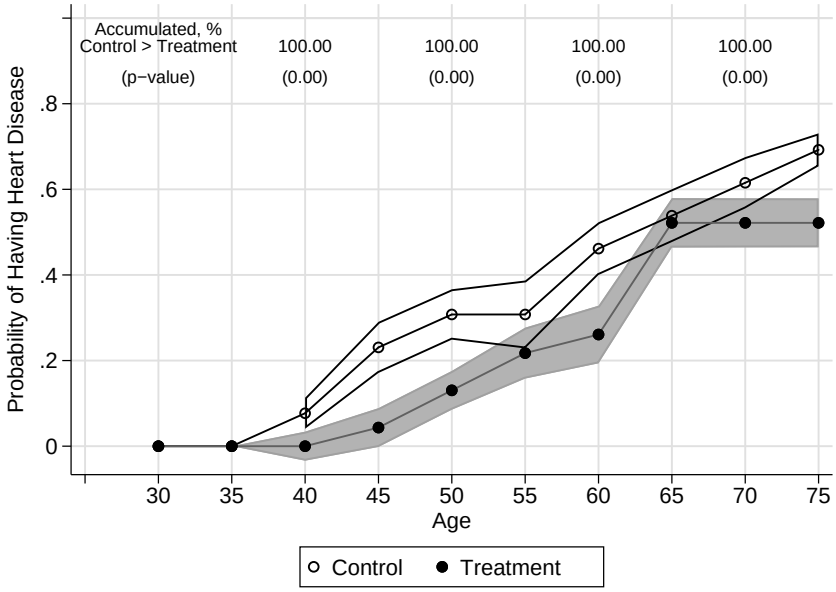
The combining functions for stroke are significant over the entire lifecycle for both genders in Figure 7. Treatment-group females have slightly lower incidence rates than control-group females throughout the entire life-cycle, in both the main simulation and all bootstrap simulations. All combining functions are at the 100% level and p-values are zero. Similarly, treatment-group males have consistently lower incidence rates than the control group. There are no cases of male stroke until age 40, after which treatment-group males have much lower incidence rates until age 65. However, age 50 and 60 combining functions have large p-values, implying larger mid-life variation in male stroke projection. The male treatment group rate

Figure 5: Life-cycle Trajectories of Heart Disease by Gender and Treatment Group

(a) Female



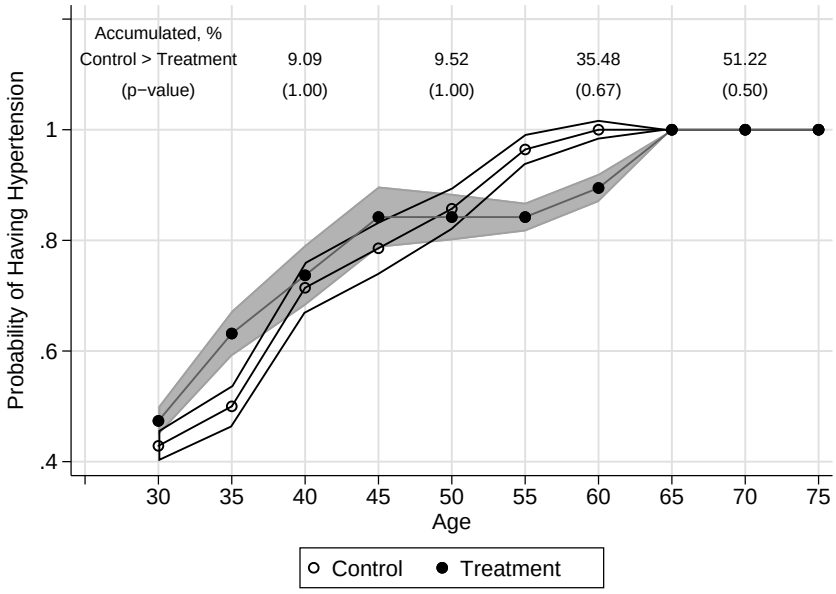
(b) Male



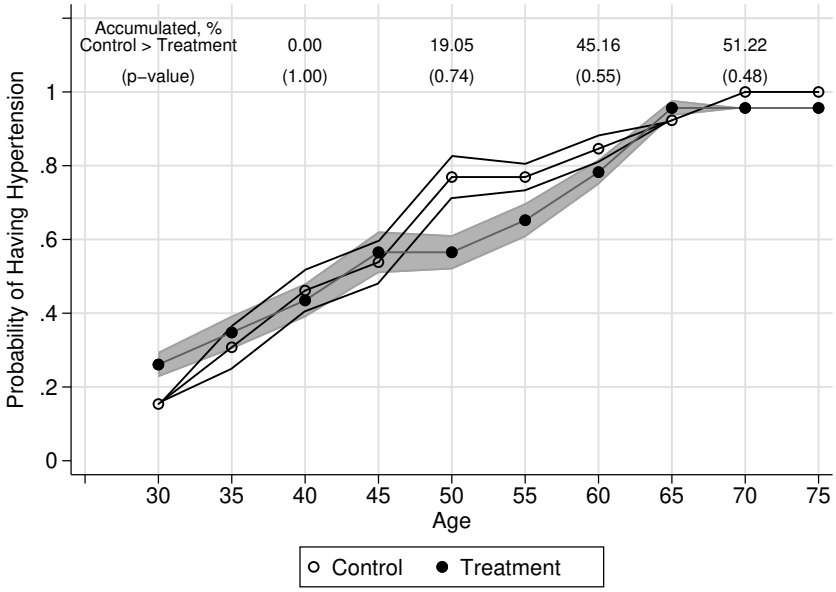
Note: The figures plot the proportion of treatment and control subjects who either have the disease or die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual contracts the disease or dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect as well as standard errors. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation, which we also use to calculate standard errors. Combining functions are highly significant except for females at age 70.

Figure 6: Life-cycle Trajectories of Hypertension by Gender and Treatment Group

(a) Female



(b) Male



Note: The figures plot the proportion of treatment and control subjects who either have the disease or die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual contracts the disease or dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect as well as standard errors. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation, which we also use to calculate standard errors. Even though there are adversely significant combining functions for hypertension, they become insignifcant beyond ages 60 and and 50 for females and males, respectively.

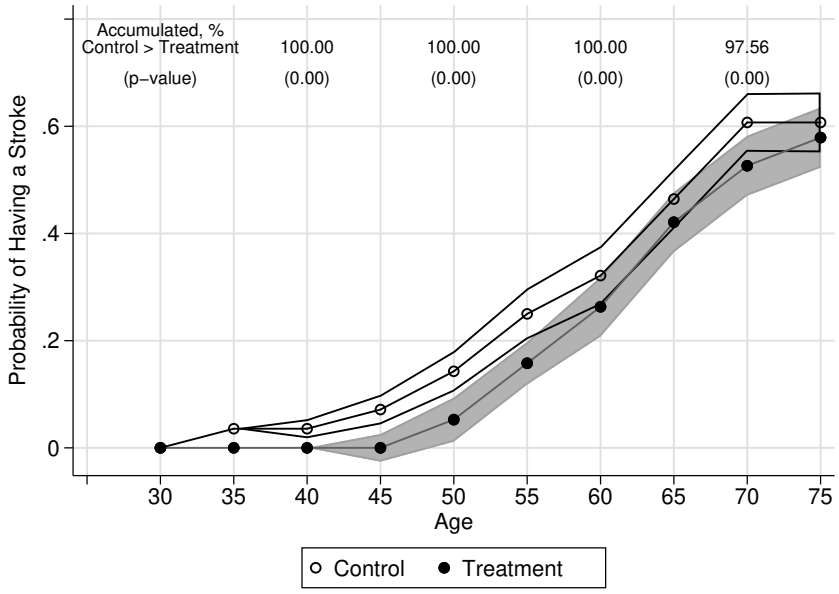
plateaus at 30% by age 65, while that of the male control group increases past 50%. The age-70 combining function becomes significant. There are highly significant combining functions for mortality across the entire life-cycle for both genders (see Figure 8).

There are no deaths until age 50 and 45 for females and males, respectively. The female treatment group has a slightly lower mortality rate than the control group over the remainder of the life-cycle. On the other hand, male mortality rates diverge over time between treatment and control. Control males are nearly four times more likely to die than treatment males by age 75. All combining functions are at the 100% level and have high levels of statistical significance, indicating similar mortality trajectories across bootstrap simulations. Mortality is an informative indicator of the aggregate effect on all of the health conditions. The prevalence of each chronic disease increases the risk of mortality. The strength and precision of the treatment-control difference across the life cycle, as well as the gender difference in the treatment effects, summarize the results in this section.

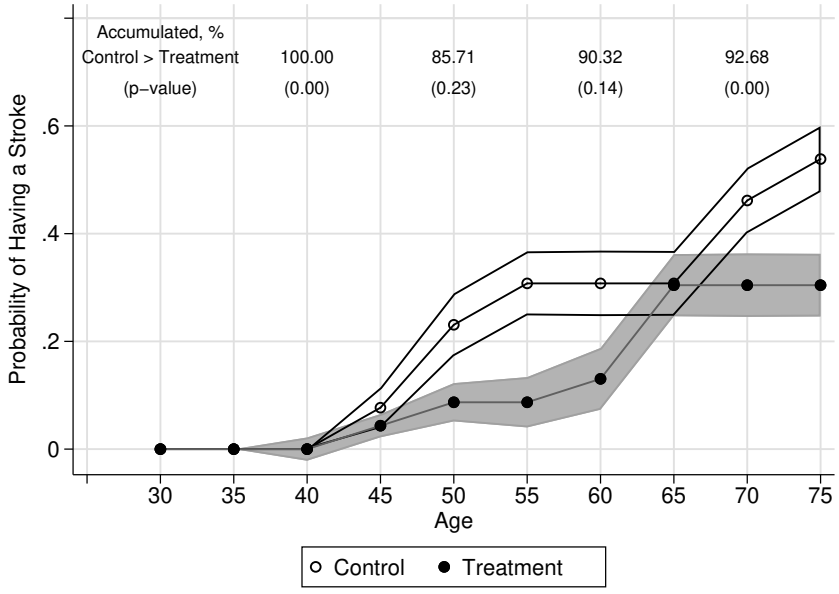
We summarize the impact of ABC/CARE on life-cycle health using quality-adjusted life years (QALYs). A QALY reweights a year of life to adjust it for disease burden. We assign a value of $150,000 (2014 USD) to each year of life. A QALY of $150,000 denotes the value of a year of life in the absence of disease (perfect health). The value of a QALY for an individual in a given year is smaller than $150,000 when there is disease: worse health conditions imply a lower quality of life. A QALY is zero at death. Table 7 presents the net present value of treatment gains in QALYs based on predicted life-cycle health trajectories. The program cost per child is $92,570 (2014 USD). Health benefits alone for treatment-group males, which total $106,218 (2014 USD), exceed the cost of the entire program and generate positive returns. By this measure alone, female health gains pay for nearly half of the program cost.

Figure 7: Life-cycle Trajectories of Stroke by Gender and Treatment Group

(a) Female



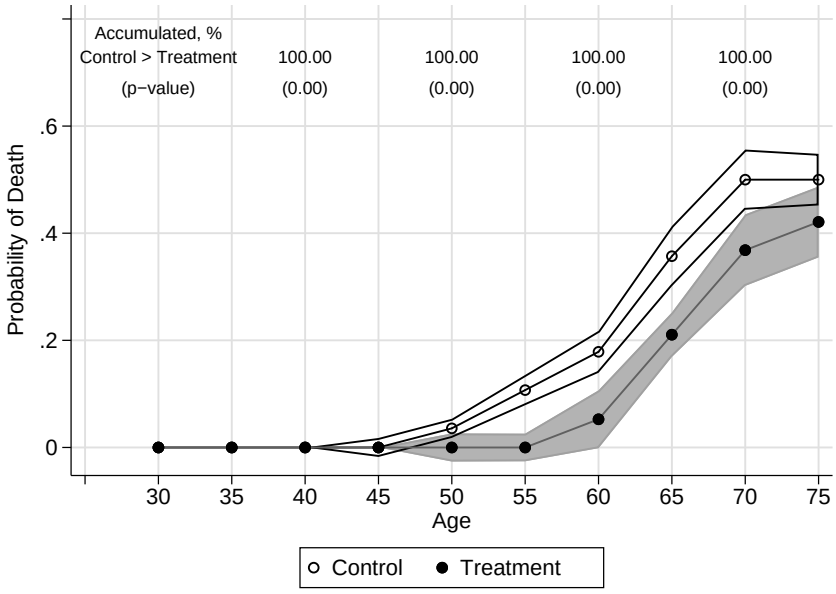
(b) Male



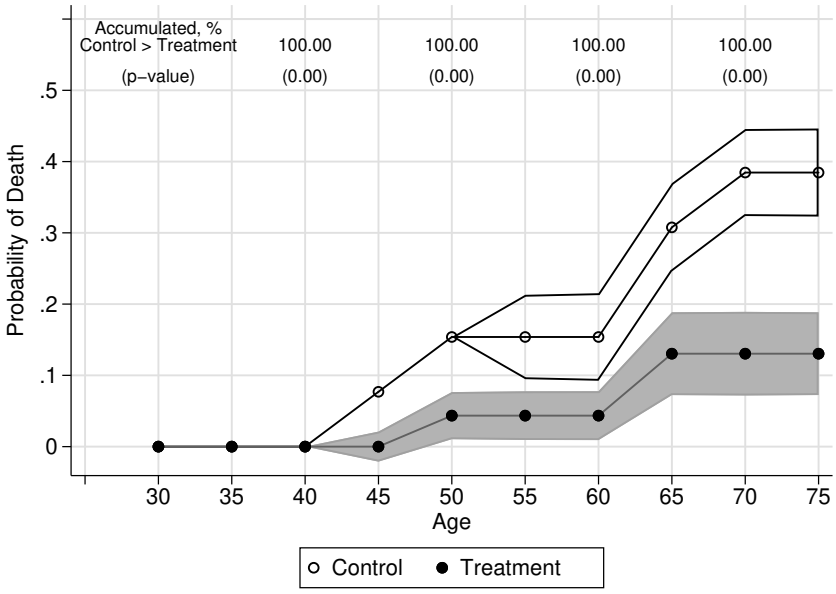
Note: The figures plot the proportion of treatment and control subjects who either have the disease or die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual contracts the disease or dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect as well as standard errors. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation, which we also use to calculate standard errors. Combining functions are highly significant throughout the life-cycle for females, and at the beginning and end of the life-cycle for males.

Figure 8: Life-cycle Trajectories of Mortality by Gender and Treatment Group

(a) Female



(b) Male



Note: The figures plot the proportion of treatment and control subjects who die over the life-cycle by 5-year bins. The individual-level 5-year bin takes the value 1 if an individual dies in any year during the 5-year window, and 0 otherwise. The profile in the figure plots the treatment and control averages of the individual-level bins. We display the fraction of years from age 30 to ages 40, 50, 60, and 70 with a positive treatment effect. One-sided p-values are calculated with a null hypothesis of 50% using 1000 bootstraps of the microsimulation. All combining functions are highly significant regardless of gender.

Table 7: ABC/CARE Impact on Life-cycle Health QALYs and DALYs

	QALYs (2014 USD)	DALYs
Females	42,102	-0.050
Males	106,218	-3.824

Note: This table presents two summaries of the life-cycle health impacts of ABC/CARE. The net present value of the average gain in quality-adjusted life years (QALYs) and the average gain in years lost due to disease burden, disability-adjusted life years (DALYs). For reference, the per-child average total cost of the program is $92,570 (2014 USD).

An alternative measure that does not depend on arbitrary monetary values for life is disability-adjusted life years (DALYs). QALYs are weighted monetary values, adjusted for disease burden. DALYs are weighted years of life, adjusted for disease burden The fraction of the year lost is summed across the life cycle to calculate the number of years lost due to disease burden. A QALY is zero at death. A DALY is one at death. We calculate the average gain in DALYs using a horizon of one hundred years of life. For men, almost four years of life is lost due to disease burden. For females, the decrease in lifespan is more moderate. Table 7 is consistent with our life-cycle predictions. We find a persistent life-cycle treatment effect for the majority of chronic conditions in both genders. The magnitude of treatment effects is larger for males in multiple disease outcomes across the life-cycle and, therefore, the life-cycle gains in QALYs and DALYs are consistent with this finding.

5 Conclusion

A well-established literature documents that high-quality early education programs benefit children in terms of test scores, schooling, crime, and adult earnings, especially those who are disadvantaged (Cunha et al., 2006; Currie and Almond, 2011; Elango et al., 2016). This evidence relies primarily on short-term gains on test scores or on adult outcomes such as crime or labor income. This paper breaks new ground by looking at the lifetime post-program benefits on health. There are substantial lifetime effects on many components of

health: cancer, heart disease, stroke, and mortality. Treatment effects are larger for males consistent with previous findings in the literature (Elango et al., 2016).

The estimated treatment effects reported here are surprisingly large, especially because the program did not target participant health, although it did subsidize visits to the doctors (but not medications) for participants for the first year of the program. Besides providing early health care, the program boosted cognitive and social skills which are known contributors to lifetime health (see Borghans et al., 2008; Campbell et al., 2014). Elsewhere (García et al., 2020) we use these estimates as part of a comprehensive cost-benefit analysis of the full range of outcomes studied in the ABC experiment.

References

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives 25*(3), 153–172.

Benjamin, E. J., M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, et al. (2017). Heart disease and stroke statistics – 2017 update: A report from the American Heart Association. *Circulation 135*(10), e146–e603.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Briggs, A., M. Sculpher, and K. Claxton (2006). *Decision modelling for health economic evaluation*. Oxford University Press.

Campbell, F. A., G. Conti, J. J. Heckman, S. H. Moon, and R. Pinto (2013, March). The effects of early intervention on human development and social outcomes: Provisional evidence from ABC and CARE. University of Chicago, Department of Economics.

Campbell, F. A., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. P. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science 343*(6178), 1478–1485.

Collins, A., B. D. Goodson, J. Luallen, A. R. Fountain, A. Checkoway, and Abt Associates Inc. (2010, June). Evaluation of child care subsidy strategies: Massachusetts Family Child Care study. Technical Report OPRE 2011-1, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

Courtemanche, C., J. C. Pinkston, and J. Stewart (2015, December). Adjusting body mass for measurement error with invalid validation data. *Economics & Human Biology 19*, 275–293.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Currie, J. and D. Almond (2011). Human capital development before age five. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4B, Chapter 15, pp. 1315–1486. Amsterdam: Elsevier Science B. V.

Educare (2014). A national research agenda for early education. Technical report, Educare Learning Network Research & Evaluation Committee, Chicago, IL.

Elango, S., , J. L. García, J. J. Heckman, and A. Hojman (2016). Early childhood education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 2, Chapter 4, pp. 235–297. Chicago: University of Chicago Press.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the life-cycle benefits of a prototypical early childhood program. Forthcoming at the *Journal of Political Economy*.

García, J. L., J. J. Heckman, and A. L. Ziff (2018). Gender differences in the benefits of an influential early childhood program. *European Economics Review 109*, 9–22.

Goldman, D. P., D. Ermini Leaf, and B. Tysinger (2016). The Future Americans Model: Technical Documentation. Technical report, Roybal Center for Health Policy Simulation, University of Southern California.

Havnes, T. and M. Mogstad (2011, May). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy 3*(2), 97–129.

Healthy Child Manitoba (2015, April). Starting early, starting strong: A guide for play-based early learning in Manitoba: Birth to six. Technical report, Healthy Child Manitoba, Winnipeg, Manitoba.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010, July). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Jensen, B. and M. Nielsen (2016). Abecedarian programme, within an innovative implementation framework (APIIF). A pilot study. Website, http://pure.au.dk/portal/en/projects/abecedarian-programme-within-an-innovative-implementation-framework-apiif-a-pilot-study(8d4da9a9-f2ff-44db-87cc-1278d1e7006c).html (Accessed 8/1/2016).

Johnson, J. H. (2003). The Changing Face of Poverty in North Carolina. *Popular Government 68*(3), 14–24.

Muralidharan, K., M. Romero, and K. Wuthrich (2019). Factorial designs, model selection, and (incorrect) inference in randomized experiments. Unpublished manuscript, University of California at San Diego.

Ramey, S. L., C. T. Ramey, and R. G. Lanzi (2014). Interventions for students from impoverished environments. In J. T. Mascolo, V. C. Alfonso, and D. P. Flanagan (Eds.), *Essentials of Planning, Selecting and Tailoring Interventions for Unique Learners*, pp. 415–48. Hoboken, NJ: John Wiley & Sons, Inc.

Richardson, D. B., S. R. Cole, H. Chu, and B. Langholz (2011). Lagging exposure information in cumulative exposure-response analyses. *American journal of epidemiology 174*(12), 1416–1422.

Sabatier, F., P. Darmon, B. Hugel, V. Combes, M. Sanmarco, J.-G. Velut, D. Arnoux, P. Charpiot, J.-M. Freyssinet, C. Oliver, et al. (2002). Type 1 and type 2 diabetic patients display different patterns of cellular microparticles. *Diabetes 51*(9), 2840–2845.

Schneider, B. and S.-K. McDonald (Eds.) (2007). *Scale-Up in Education*, Volume 1: Ideas in Principle. Lanham, Maryland: Rowman & Littlefield Publishers.

Scott, L. J., K. L. Mohlke, L. L. Bonnycastle, C. J. Willer, Y. Li, W. L. Duren, M. R. Erdos, H. M. Stringham, P. S. Chines, A. U. Jackson, et al. (2007). A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*.

Scull, J., J. Hattie, J. Page, J. Sparling, C. Tayler, A. King, V. Nossar, A. Piers-Blundell, and C. Ramey (2015). Building a bridge into preschool in remote Northern Territory communities.

Sousa, A. G., L. Selvatici, J. E. Krieger, and A. C. Pereira (2011). Association between genetics of diabetes, coronary artery disease, and macrovascular complications: Exploring a common ground hypothesis. *The Review of Diabetic Studies 8*(2), 230.

Sparling, J. (2010). Highlights of research findings from the Abecedarian studies. Technical report, Teaching Strategies, Inc., Center on Health and Education, Georgetown University, and FPG Child Development Institute, University of North Carolina at Chapel Hill, Bethesda, MD.

Spiker, D., C. W. Haynes, and R. T. Gross (Eds.) (1997). *Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program.* Redwood City, CA: Stanford University Press.

van Baal, P. and H. Boshuizen (2019). Modeling chronic diseases in relation to risk factors. In *Oxford Research Encyclopedia of Economics and Finance*.

Yazejian, N. and D. M. Bryant (2012). Educare implementation study findings. Technical report, Frank Porter Graham Child Development Institute, Chapel Hill, NC.

Zeggini, E., L. J. Scott, R. Saxena, B. F. Voight, J. L. Marchini, T. Hu, P. I. de Bakker, G. R. Abecasis, P. Almgren, G. Andersen, et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics 40*(5), 638.