Dissertation

A Lot on their Shoulders

Predictive Modelling Approaches for Addressing Shrinking Working Populations in Select Countries

Ifeanyi Edochie

This document was submitted as a dissertation in August 2020 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Jeanne Ringel (Chair), Osonde Osoba, and Troy Smith.

Partial funding for this dissertation was provided by the Pardee Initiative for Global Human Progress.

RAND PARDEE RAND GRADUATE SCHOOL

For more information on this publication, visit http://www.rand.org/t/RGSDA2236-1.

Published 2022 by the RAND Corporation, Santa Monica, Calif.

Abstract

In many countries, age dependency has been on the rise placing pressures on the working age populations to provide for the socio-economic needs of the young (ages 0 -16) and elderly (ages 65+). This dissertation attempts to address the policy problems generated by age dependency problems using predictive analytic methods.

The first paper micro-simulates the effects of a cash transfer program (both conditional and unconditional) of the same individual amount on educational enrollment in basic education for children aged (7-15) and the intensity of poverty for Nigeria. This study shows that increasing program coverage and providing smaller conditional transfers per child in school is more effective at increasing educational enrollment more than increasing the transfer amount. The second study projects the demand for dementia care using a novel microsimulation approach in the United States. The results suggest that the age distribution of the population is and will continue to be an important factor driving dementia prevalence. The final study uses a machine algorithm to predict dementia status across individuals without using any measures of cognition. We show that a modelling solution like this may be used as a first step to eliminate individuals that might not need the more traditional brain scans that are usually more expensive. Therefore, there may be cost savings in the future attributed to this approach but further methodological research is needed.

Governments in both the Western countries and Africa have significant upcoming challenges to deal with as age dynamics evolve over the next 50 years in their respective countries. As cash transfer programs that seek to increase educational enrollment become more prevalent, African governments need to define clearer child labor policy. African development, in part, is dependent on the ability of policymakers to be create the environment for academic innovation as there are no clear cut solutions to deal with the large proportion of young dependants that require education but come from poor households. In the countries with old age dependency, policymakers will have to grapple with the rising prevalence of age-old diseases amongst the overall population. This dissertation is an attempt to begin supporting this line research.

Contents

Acknowledgments

I am deeply grateful to all my parents, teachers and classmates whose support and advice has been instrumental through the dissertation process.

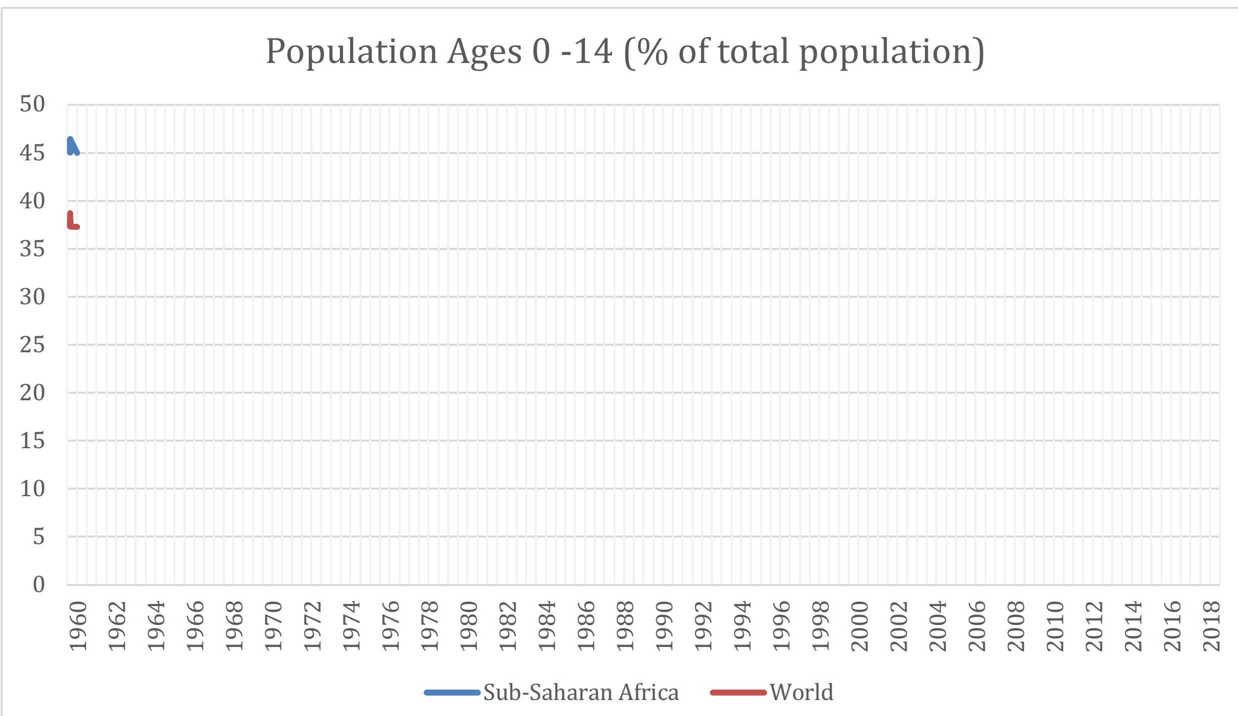
Abbreviations

CCT	Conditional Cash Transfers
UCT	Unconditional Cash Transfers

Introduction

In many countries, age dependency has been on the rise placing pressures on the working age populations to provide for the economic needs of the young (ages 0 -16) and elderly (ages 65+). The dependency ratio in a country or region is the number of individuals outside the working age (i.e. those ages 0 to 16 and 65+) to those in the working age population. In African countries, high fertility rates have led to the youth bulge so that large proportions of these countries are comprised of individuals between the ages of 0 and 16.

Figure A1

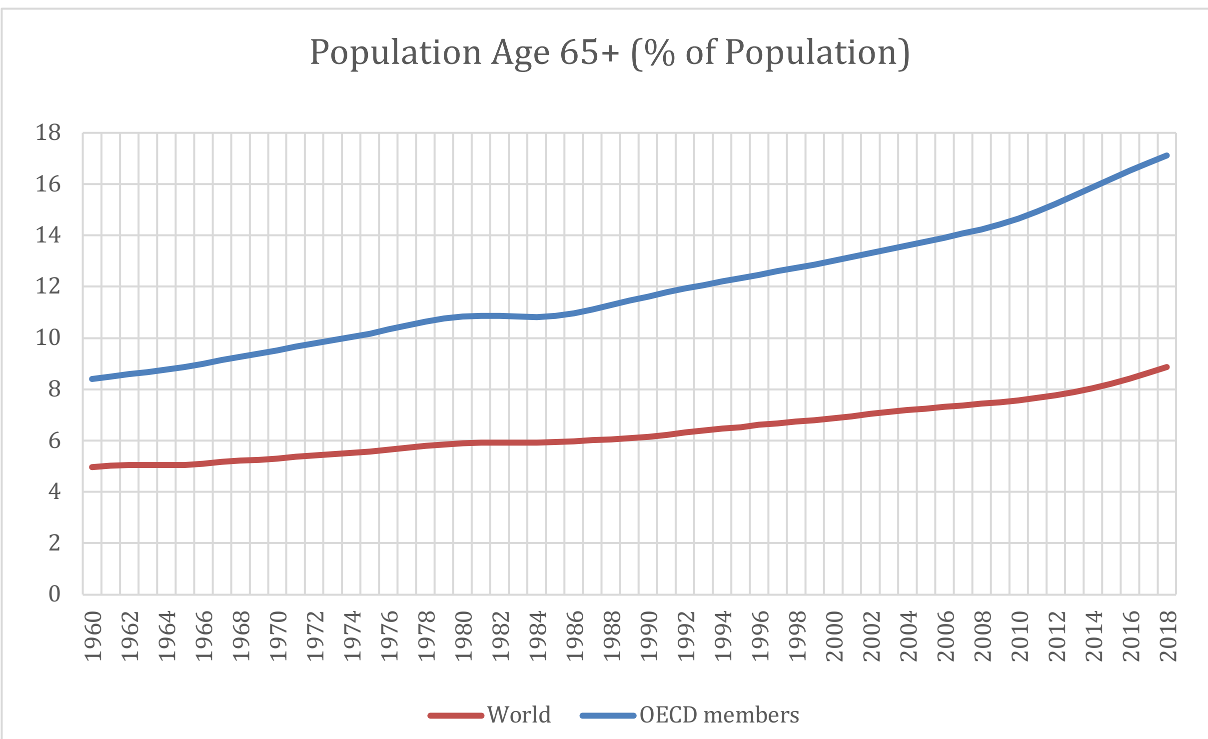


Source: World Development Indicators

The chart above shows a proportion of youth (ages 0 to 14) for all countries (world) as well as for all Sub-Saharan African countries. The youth rate in African countries has remained relatively constant and high at between 42-46% from 1960 to 2018 while the global rates have fallen from 37% to about 25% in the same time period. Consequently, smaller working populations have to provide for the economic needs of large youth populations as well as the shrinking share of elderly as well. This includes the household's decision of what occupational choices are made for children (i.e. whether to attend school, work, do both etc). Families often make decisions of how much to invest in their children's schooling as well as how much and what type of employment children engage in. This is important especially in sub-Saharan Africa due to the high prevalence of low wage informal work as well as weak social protection systems that fail to subsidize consumption for poorer households. Therefore, working age population in African countries as well as other developing countries are ill-equipped to support the large population of children in their countries.

On the other hand, high-income countries are riddled with low fertility rates and advancements in medical technology have led to prolonged life as well as aging populations.

Figure A2



Source: World Development Indicators

From the chart above, it is evident the proportion of the population that is elderly is significantly higher in the OECD countries. In contrast to the youth bulge, this creates a different economic pressure on the working population requiring them to support services for those over the ages of 65. As populations age, diseases such as dementia associated with the elderly population become more prevalent in society. Public health care systems become more stretched for resources to take care of their elderly. Therefore, the ability of aging countries to provide elderly care will grow in importance as the effects of an aging population becomes more prevalent in a country with a low fertility rate.

Predictive modelling approaches could play a significant role in highlighting and mitigating the future challenges introduced by the specific population age structures previously described. A Microsimulation modelling approach mimics the behavior of units (could be individuals, public and private sector systems etc) towards evaluating the effect of phenomena such as intervention of policy outcomes of interest. For instance, microsimulation models have been used to represent the US health care system towards measuring the effects of certain regime's health policies on premiums and national coverage. The approach typically brings data together from multiple sources such that when decision rules are applied to the model it can be used to compare effects of counterfactual policy scenarios to the status quo. The advantage of simulation techniques lie in their ability to simulate non-existing but realistic scenarios as well as measure the effects of policies on specific sub-groups that may often by marginalized by policies. On another hand, machine learning approaches use classification algorithms to make policy relevant predictions. The approach is flexible across a wide variety of contexts. For instance, machine learning algorithms can be used to predict dementia status towards creating screening tools. They have also been used to identify the extreme poor (by the $1.90 poverty line) in social protection programs within developing countries.

This dissertation seeks to address these two related problems using predictive modelling approaches. The first paper assesses the use of (un)conditional cash transfer programs to reduce poverty in households and increase educational enrollment in a country like Nigeria. This study is thus concerned with supporting the working age population within the country in dealing with the youth bulge. The analysis uses a microsimulation modelling approach to assess the likely poverty and enrollment effects of different cash transfer regimes i.e. estimating the income effect (on school enrollment) and applying it to a representative sample of the Nigerian population. In the second study, we project the demand for dementia care in the United States up till 2060. We develop a novel microsimulation approach that simulates dementia incidence unto an existing model of the US population between 1910 and 2060 containing demographic information including familial structure. This is important for estimating the demand for dementia care as well as the potential supply of family caregivers. This is an important preliminary step for policymakers making futuristic plans for elderly care. In the final study, we are concerned with the process of disease detection for the specific case of dementia. The most sophisticated studies apply machine learning algorithms to classify brain scans as either being owned by a dementia-positive or dementia-free patient. We apply a machine learning algorithm to the RAND Health and Retirement Study (HRS) to classify elderly respondents by dementia status using only demographics characteristics, activities of daily living measures as well as presence of pre-existing conditions. This is important because the economically disadvantaged are less likely to have insurance which makes the ability of obtain a diagnosis more difficult. A machine learning algorithm that predicts dementia status based on non-cognitive variables significantly reduces the cost of detection, thereby making care more accessible to the poor.

Chapter 1: Micro-simulating the Education Enrollment Effects of Conditional and Unconditional Cash Transfers in Nigeria

Ifeanyi Edochie

December 19th, 2019

1.0 Introduction

Social safety net programs, specifically cash transfers, have grown in prominence as the policy tool of choice for poverty reduction and other socio-economic policy goals such as maximizing school attendance/enrollment amongst children, health care utilization etc. Cash transfers are generally of two kinds: conditional or unconditional. Conditional cash transfer (CCT) are monetary benefits given to households by government or an enforcing body under the conditions that certain measurable "co-responsibilities" are met by the recipients. For instance, a government in country/region may choose to provide transfers to poor and vulnerable households[1] on the condition that they send their children of a specific age to school. On the other hand, unconditional cash transfers (UCTs) are provided without expectation that recipients will perform certain actions, in theory. However, in practice, UCTs often have "soft conditions" i.e. a program might provide funds while trying to encourage individuals to attend school or utilize primary health care facilities. While providers of UCTs may hope for positive outcomes among recipients due to the additional resources provided, they do not directly incentivize a specific change in behavior in the same way a CCT would. In this study, I quantify the tradeoffs between improvements in school enrollment and poverty reduction due to education-based CCTs vs UCTs.

Conditional and unconditional cash transfers possess certain trade-offs as previously mentioned. Imagine, a government has a budget to provide transfers to poor households within a specific region. If the body implements a UCT providing transfer sum, t, to each poor household, the

[1] Poor and vulnerable households are based on a measure of household earnings or consumptions typically determined by a proxy means testing approach.

UCT in theory will reduce population-wide poverty more than a CCT that provides the same, t, to each poor household that also requires children to enroll in school to be transfer eligible. This is obvious because the same, t, provided is now available to fewer households in the CCT case than in the UCT. However, the education based-CCT should, in theory, increase school enrollment and attendance better than a UCT simply because households in a CCT program have to attend school to receive any transfers. Modelling these trade-offs within a country's specific context is important since practical concerns often constrain program implementation. For instance, policymakers in a country may perceive it impossible to implement a nationwide CCT if the cost of monitoring compliance is too high. In that case, the country might choose to keep the same individual transfer amount, t, the same but instead exhaust the program budget over a shorter timeframe. These program implementation concerns are important as African governments have been implementing these nationwide poverty eradication projects with an eye on meeting poverty targets of the Sustainable Development Goals for 2030 as well as Africa's Sustainability Agenda for 2063.

This analysis simulates the effects of a cash transfer (both conditional and unconditional) of the same individual amount on educational enrollment and the intensity of poverty for a country. Therefore, this study seeks to estimate overall transfer costs to assist a country in deciding when the cost of monitoring an education-based CCT may be worth absorbing rather than running a UCT that might not, in theory, yield lesser education enrollment improvement for the same per household transfer amounts.

1.1 The Cash Transfer Literature

The cash transfer literature can be broken up into 2 major strands.

(i) Ex-Post impact evaluations which use specific program intervention data to measure causal effects. These include standard evaluations of the impacts of providing monetary transfers on outcomes such as education, health & nutrition, labor outcomes (more common with public works programs), migration, fertility as well as impacts on intra-household allocation (behavioral effects).

(ii) Ex-Ante studies are far less common but are necessary to model the effects of different program features on policy outcomes of interest in a manner that seeks to directly improve a prospective program's design features. We propose to do this within a microsimulation framework using a representative sample of the Nigerian population (General Household Survey 2015/2016). The purpose of this study to improve program intervention design based on the likely effects of changing certain program features on educational enrollment and program costs.

The most commonly evaluated programs have been Bolsa Escola/Familia in Brazil and Mexico's Progressa. To a lesser extent, other programs in Cambodia, Venezuela and Chile have also been reviewed [1, 2]. Even fewer programs in African countries have been studied, which highlights the importance of this study's focus on Nigeria. This is the largest African country, resident to about 13% of the continent's population with about 47%[3] of its inhabitants earning less than $1.90 a day. The recent rebasing of the country's GDP made Nigeria the largest economy in Africa. With Nigeria as well as other African countries have been implementing cash transfer programs, this study aims at using simulation techniques to show different educational enrollment increases created by choosing amongst UCT or a CCT programs subject to program costs. This helps us quantify the tradeoffs between education enrollment and poverty from a UCT and a CCT providing the same transfer amount. In addition, we can compare overall transfer costs to the government under each program to shed light on whether the government may invest in monitoring a CCT, given sizably larger costs of a UCT program

1.1.1 The Ex-Post Literature in brief

[4] reviews the existing evidence on the long-term impacts of conditional cash transfer programs on education and future job prospects of children from poor families. Some of the most prominent studies measuring impacts of CCTs on schooling are on programs from Latin America especially Bolsa Escola [5, 6]. In addition, similar programs in African countries are on the rise. The program effects on school attendance in the literature range from near 0% to over 20%. [2] report on the estimated effects on school attendance from Chile's Solidario (7.5%), Ecuador's Bono de Desarrollo Humano (10.3%), Honduras' Programa de Asignacion Familiar (3.3%), Jamaica's Program of Advancement of Health and Education (0.5%) and Nicaragua's Red de Proteccion Social (12.8%). However, program effects may vary within specific sub-populations. [1] in studying Mexico's Opportunidades finds an effect of 1.9% (statistically insignificant) for children in grades 0-5 and grades 7-9 at 0.6% (statistically insignificant). Another study[7]

measures the effect of Kenya's Cash Transfer for Orphans and Vulnerable Children (OVC) on human capital. The study finds that "among households living over 2 kilometers from a primary school, the treatment effect on current enrollment is 19 percentage points higher and 6 percentage points higher for each unit increase in the primary school cost index".

1.1.2 The Ex-Ante Literature in brief

Model based approaches have also been used to measure the ex-ante impact of conditional cash transfer programs on poverty [8, 9]. Most methods typically estimate the effect of a cash transfer on educational enrollment occurs through an income effect. The advantage of these model-based approaches is in the ability to simulate counterfactual scenarios. However, in most cases, these models are unable to guarantee casual effects. [8] explains that this is less than ideal, however, this method, allows for the computing of long term effects in a way that most ex-post studies might not be able to.

[10] investigates the long-term general equilibrium impacts of CCTs on school attainment and child labor outcomes. The results suggest the Bolsa Escola program could reduce the share of children working from 22% to 17% in the near term while showing that in the longer term the share of children completing primary education could rise from 52% to 90%. Many of the ex-ante studies have been shown to approximate the effect sizes seen in the ex-post studies lending to their viability as an effective tool for planning interventions [11].

While the theoretical tradeoffs between CCTs and UCTs have been discussed in general, we find no studies that sought to quantify them within the specific context of a country. In 2017, Nigeria chose to implement a UCT in place of an education-based CCT as the government deemed a CCT impossible to monitor within the country. We attempt to shed more light on the effects of this decision by using a microsimulation approach to:

1) compare the tradeoffs between education enrollment and poverty from a UCT and a CCT providing the same transfer amount.

2) compare overall transfer costs to the government under each program to shed light on whether the government may invest in monitoring a CCT, given sizably larger costs of a UCT program.

1.2 Context: Nigeria's Conditional Cash Transfer Program

In 2017, the Federal Government of Nigeria (FGN) sought to implement CCT programs run at the state level to provide funds to poor households if they sent their children to school. However, the states were unable to monitor these programs. In the original plan, poor households (based on proxy-means testing[2]) would be offered 5000 Naira (NGN) if they sent their children between the ages of 7 and 15 to school. It is unclear whether the funds would be provided per child or per household (our analysis accounts for this uncertainty). The FGN decided to implement only the unconditional transfers since monitoring education-based conditional transfers was deemed to be too difficult to manage at the national level. Assuming perfect targeting, the overall transfer costs (i.e. the product of transfer sum and number of recipients) is higher for a UCT than for a CCT because conditional transfer program eligibility is higher to reach than a UCT. However, if a UCT costs more without providing sufficient increases to educational enrollment, how much more should a government be willing to invest in monitoring a CCT given its benefits to educational enrollment?

In attempting to address these issues, we simulate the effect of cash transfers on educational enrollment and the intensity of poverty in Nigeria. We compare the likely poverty reduction - educational enrollment effect trade-offs emanating from implementing conditional or unconditional programs. We also compare the overall transfer costs of both program types under multiple designs assuming they provide the same transfer sum. The results of the analysis show that for the same transfer amount ($14 per household/month), around 60-70% of transfer eligible children from poor families will enroll in school as a result of the conditional cash transfer. This rate falls to approximately 3% for a UCT offering the same transfer amount to ALL poor households. On the other hand, a UCT reduces distance on average between each poor household's income and the poverty line ($1.90 a day) from 40.14% to between 13.73% to 33.54% (depending on the UCT program design) while a CCT only reduces poverty to between 35.7% and 37.8% (also depending on the CCT program design). Using the ratio of UCT to CCT overall transfer cost, we estimate a decision rule at which a government should be willing to

[2] Proxy-means testing uses regression or machine learning models to estimate what households are poor (i.e. cash transfer eligible) as a function of observable household characteristics. This makes it easy for transfer program workers to apply the model coefficients to characteristics specific to a household towards estimating whether or not the household is cash transfer eligible.

absorb the cost of monitoring. We find that certain CCT programs cost less UCT programs, therefore Nigerian program administrators have some leeway under specific conditions that allow for investing in the monitoring of their conditional cash transfers.

1.3 Model: A System for Modeling and Simulating a Cash Transfer Program in Nigeria

An ex-ante analysis measuring the potential effects of a cash transfer policy on student enrollment is difficult for several reasons. Measuring income in Nigeria is problematic; only about 6% of workers are salaried, while the majority work within the informal sector where income is irregular and often difficult to capture accurately within a survey. Child incomes/earning potential are unobservable as most (if not all) Nigerian children work in a household enterprise (either farm or non-farm) as opposed to salaried work where pay can be documented. However, a measure of child earning potential is necessary for estimating the decision to attend school. The occupational decision-making process for a child is a complex process which is intertemporal i.e. many individual and household factors determine whether a child attends school and during what periods. Consequently, sophisticated dynamic models would be necessary for any robust results on the likely income effects on educational enrollment. However, non-causal reduced form models have been widely used in this field [8, 12] and been shown to provide results that often match ex-post study estimates [11].

Our model makes a few important simplifying assumptions. The child's decision is made by the head of the household in a unitary fashion which is a simplification of the reality but we believe is reasonable to make approximations of the likely income effects a cash transfer program might have. The program is implemented at the same time nationwide and each family decides on the occupational status of their children the moment they receive a transfer or offer (in the case of the conditional transfer). As a result, we can estimate a static choice model. We also do not fully address the obvious endogeneity of our income measure as a determinant of school choice. In this vein, we follow the lead of standard literature in the field that has computed "order of magnitude effects" [8].

1.3.1 Theoretical Model

We adapt an existing theoretical model [8] to fit the socio-economic realities of an African country like Nigeria. The modelling approach is predicated on measuring an income effect, i.e. a cash transfer increases family household income. If cash transfers are to be effective a change in household income should increase the probability a child attends school. The model assumes every household makes the work-choice decision of each individual in the family in a unitary model, typically the head of the household dictates what each member participates in and to what extent. Each household's decision set for child, i, S_i, is given by

$$k \subseteq \{0, 1, 2\}$$

where S_i represents the occupational choice made for a child in household i. $S_i = 0$ if a child is not in school. $S_i = 1$ when the child attends school and works within the labor market. $S_i = 2$ refers to the state of only attending school. Each child's occupational choice S_i follows the standard utility-maximizing multinomial logit regime as follows:

$$(1) \quad S_i = k \quad iff \quad S_k(A_i, X_i, H_i, \mathrm{Y}_{-\mathrm{i}}; y_i + h_{ik}) + v_{ik} \ > \ S_j(A_i, X_i, H_i, Y_{-i}; y_i + h_{ij}) + v_{ij} \quad \forall j \neq k$$

where the household head makes a decision based on $S_k()$, a latent function reflecting the net utility of choosing alternative, k. A_i is the child's age, X_i and H_i refers to the set of child (gender, birth position, etc.) and household characteristics (household head schooling years, age of parents, size, type of household business run etc). Y_{-i} measures household monthly earnings less the contribution of the child while y_i measures the monetary contribution of each child. v_{ik} and v_{ij} are the standard random variables capturing unobserved heterogeneity of the observed S_i choice.

We specify a linear utility function by combining all child and household variables into a vector Z_i.

$$(2) \quad U_i(j) = S_j(A_i, X_i, H_i; Y_{-i} + y_i) + v_{ij} = Z_i.\gamma_j + (Y_{-i} + y_i)\alpha_j + v_{ij}$$

where γ & α are the parameters representing the order of magnitude effects in the model.

1.3.2 Estimating Potential Child Wages

Children working in the labor market in Nigeria largely work within a family owned business. Consequently, formal pay does not apply to them. As a result, we compute child wages from hours worked by the children in $S_i = 0{,}1$ under the assumption of constant marginal productivity of each household family member i.e. child wages are directly proportional to their share of family hours worked. The discrete choice model requires the estimation of potential expected earnings of a child including those who do not work outside the household. Therefore, we assume that families consider the potential earnings of a child as part of the decision making since the opportunity cost of school attendance is the lost wages of the child. For the purposes of this analysis, we simply estimate a simple Becker-Mincer equation of the following form:

$$(3) \quad log(w_i) = Z_i\theta + m * Ind(D_i = 1) + u_i$$

where $Ind(\dots)$ is an indicator function where D takes the value 1 if children are performing labor market work (S_i = 0). This term accounts for the fact that individuals working in the labor market without attending school should contribute more to their families than children attending school and working in the labor market. Based on equation 3, the contribution of the child to household income varies with total family income by occupational choice as follows:

$$(4) \quad y_{i0} = Kw_i \quad for\ k = 0$$

$$y_{i1} = My_{i0} = MKw_i \quad for\ k = 1$$

$$y_{i2} = Dy_{i0} = DKw_i \quad for\ k = 2, \ \ with\ M = e^m$$

We use y_{ij} to measure the value of domestic and market child labor. For the children who are not in school, $y_{i0} = Kw_i$. Domestic income, w_i , is proportional to potential value of child labor, y_{i0}, with a proportional constant, K. For those schooling and working, $y_{i1} = My_{i0} = MKw_i$. This implies the potential value of child labor for children in both school and work, y_{i0}, is $1 - M$ the value of labor for children who don't attend school, y_{i1}. Likewise, children who school without working have a value worth $1 - D$ of the potential income of children who do not attend school. For the purposes of our analysis, we expect M to be greater than 0 and less than 1. D, however, is less than or greater than 1. Since equation 3 is only estimated with $k = 0{,}1$, it is uncertain how the potential labor value of children who would end up only in school would be valued relative to y_{i0}.

Combining equation 2 with the system of equations 4, we get the following utility function:

$$(5) \quad U_i(j) = S_j(A_i, X_i, H_i; Y_{-i} + y_{ik}) + v_{ij} = Z_i.\gamma_j + Y_{-i}\alpha_j + y_i\beta_j + v_{ij}$$

with $\beta_0 = \alpha_0 K, \beta_1 = \alpha_1 MK, \beta_2 = \alpha_2 DK$

After computing α, β, γ, potential child income and the residuals v_{ij}, the household head's choice for child i, k, is maximized according to:

$$(6) \quad k^* = Arg\ max_j[U_i(j)] \quad \forall j = 0,1,2,3$$

If the government entitles all children to a transfer amount, T, equation 5 becomes:

$$(7) \quad U_i(j) = S_j(A_i, X_i, H_i; Y_{-i} + y_{ik}) + v_{ij} = Z_i.\gamma_j + (Y_{-i} + NCTP_{ij})\alpha_j + w_i\beta_j + v_{ij}$$

$$with\ NCTP_{i0} = NCTP_{i1} = T; NCTP_{i2} = NCTP_{i3} = 0$$

This simply introduces a transfer amount, T, to the income vector for individuals who attend school in the CCT case. However, the UCT transfers T to each family regardless of occupational choice. A transfer targeting the poor would be preceded by a proxy means test and then the same conditions apply for individuals falling below a threshold. Under these assumptions, equations 6 and 7 are the full reduced form models of occupational choice of children. We can simulate the likely enrollment effects of Nigeria's National Cash Transfer Program (NCTP) transfers. In the next section, we explain the framework for identifying all parameters.

1.3.3 Simulating the Enrollment Effects of Cash Transfer Program Designs in Nigeria

A conditional transfer offers a transfer amount, t, to families that fall below a certain income threshold who send their children to school. Each state offers the equivalent of $14 (approximately) to 10% poorest families who may or may not send their children to school until their budget is spent. Each potential recipient household i will choose j that maximizes the utility from program-specific alternatives:

(9)

$$U_i(S_i = 0) = Z_i.\gamma_0 + (Y_{-i} + T)\alpha_0 + w_i\beta_0 + v_{i0} \quad if\ Y_i \leq Y^0$$

$$U_i(S_i = 0) = Z_i.\gamma_0 + Y_{-i}\alpha_0 + w_i\beta_0 + v_{i0} \quad if\ Y_i > Y^0$$

$$U_i(S_i = 1) = Z_i.\gamma_1 + (Y_{-i} + T)\alpha_1 + w_i\beta_1 + v_{i1} \quad if\ Y_i \leq Y^0$$

$$U_i(S_i = 1) = Z_i.\gamma_1 + Y_{-i}\alpha_1 + w_i\beta_1 + v_{i1} \quad if\ Y_i > Y^0$$

$$U_i(S_i = 2) = Z_i.\gamma_2 + (Y_{-i} + T)\alpha_2 + w_i\beta_2 + v_{i2}$$

1.3.4 Estimating the Discrete Occupational Choice Model

We use the well-known multinomial logit model to estimate occupational choice assuming that the $v_{ij}s$ are independently and identically distributed across sample observations. This model estimates that the probability that household i will select occupational choice k is given by:

$$(12)\ p_{ik} = \frac{e^{Z_i.\gamma_k + Y_{-i}\alpha_k + w_i.\beta_k}}{\sum_j e^{Z_i.\gamma_j + Y_{-i}.\alpha_j + w_i.\beta_j}}$$

In computing probabilities for each level j with reference level, j = 0, the values are written as follows:

$$(13)\ p_{ij} = \frac{e^{Z_i.(\gamma_j - \gamma_0) + Y_{-i}(\alpha_j - \alpha_0) + w_i.(\beta_j - \beta_0)}}{1 + \sum_j e^{Z_i.(\gamma_j - \gamma_0) + Y_{-i}(\alpha_j - \alpha_0) + w_i.(\beta_j - \beta_0)}}$$

for $j = 1,2$ and $p_{i0} = 1 - p_{i1} - p_{i2}$

The difficulty is that the multinomial logit estimation permits identifying only the differences $(\gamma_j - \gamma_0)$, $(\alpha_j - \alpha_0)$ and $(\beta_j - \beta_0)$ for $j = 1,2$. Inspection of (9) shows that a simulation of a transfer program is state contingent so that the $(\alpha_0, \alpha_1, \alpha_2)$ must be uniquely identified. Let $\widehat{\alpha_j}$ and $\widehat{\beta_j}$ be estimated coefficients from the regression model. By combining equations (4) and (5), we estimate the structural parameters as follows:

(14)

$$\alpha_1 - \alpha_0 = \widehat{\alpha_1}$$

$$\alpha_2 - \alpha_0 = \widehat{\alpha_2}$$

$$\beta_1 - \beta_0 = \widehat{\beta_1}; \quad K(M\alpha_1 - \alpha_0) = \widehat{\beta_1}$$

$$\beta_2 - \beta_0 = \widehat{\beta_2}; \quad K(D\alpha_2 - \alpha_0) = \widehat{\beta_2}$$

M can be computed from equation 3 ($M = e^m$). However, to estimate the values $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2$ *and* β_3 an assumption needs to be made about constant C. Our analysis assumes K = 1 which implies the total human capital investment of a child equals expected/potential education expenditure for children who are working but not attending school. We identify each of the coefficients as follows:

(15)

$$\alpha_1 = \frac{\widehat{\alpha_1} - \beta_1}{1 - M}; \ \alpha_2 = \widehat{\alpha_2} + \alpha_0; \ \alpha_0 = \alpha_1 - \widehat{\alpha_1}; \ D = \frac{\widehat{\beta_2} + \alpha_0}{\alpha_2};$$

1.4 Data and Descriptive Statistics

1.4.1 Description: Children and Families of Children Aged 7-15 in Nigeria

We use the Nigerian General Household Surveys (GHS) 2015/2016 to estimate equations (3) and (12) and simulate the likely effects on educational enrollment of conditional and unconditional cash transfer program designs. The survey is a sample based on around 5000 households (over 25000 individuals) representative of the Nigerian population.

Tables 1.1 presents the rates of participation in each occupational choice category for the children between ages 7 and 15. The top half presents results for the general population while bottom half provides rates for the kids in the bottom 10% poorest households in each state, i.e. the assumed transfer eligible population. It is important to note that the tables present school enrollment since the GHS is unable to track school attendance for children.

Around 78.4% of all children ages 7-15 are enrolled in school without working with 6% attending school while working and another 15.6% simply don't attend any school. Some of these children work in the labor market but we assume that all these children are working even if it's in domestic labor, i.e. households won't allow their children to be idle. The rates of school enrollment are significantly lower for children from poorer households. Only 65% of the children from the poorest households are enrolled in schooling while approximately 28% do not attend school at all. However, school enrollment appears to be a declining function of age. From the tables, school enrollment appears to be high between ages 7 and 11 as children are perceived too young to work by most families. These rates, however, drop precipitously particularly after age 11 with increasing rates of out of school children in Table 1.1. In both tables, a small class of children appear to work and go to school with a significantly higher rate of both for children from the poorest families.

Table 1.1

General Population: Occupational Status of Children Ages 7-15

	7	8	9	10	11	12	13	14	15	All
Not Attending School	15.6	17.7	14.7	15.2	11.8	14.8	15.5	16.0	19.3	15.6
Working and Schooling	3.7	3.1	4.3	6.1	5.0	7.4	9.1	8.2	8.6	6.0
Schooling without Working	80.7	79.2	81.1	78.7	83.3	77.9	75.3	75.8	72.1	78.4

Note:
Source: Nigeria General Household Survey Panel 2015/16 and author's calculations

10% Poorest All States: Occupational Status of Children Ages 7-15

	7	8	9	10	11	12	13	14	15	All
Not Attending School	25.1	26.6	23.3	22.7	28.1	22.0	38.0	25.3	44.6	27.9
Working and Schooling	6.5	5.8	7.1	4.2	8.2	7.3	5.4	8.5	15.6	7.1
Schooling without Working	68.4	67.6	69.6	73.1	63.6	70.7	56.6	66.3	39.8	65.0

Note:
Source: Nigeria General Household Survey Panel 2015/16 and author's calculations

The average child and household characteristics of the children in the Universal Basic Education (UBE) age group by occupational status can be found in table 1.3. The table shows children schooling and working in the labor market are typically older than children not attending school or only schooling without working. As expected, households that don't encourage school enrollment tend to be poorer, appear to be larger, have more children under the age of 5 (requiring child care) and are over represented within the rural communities and in the Northern parts of the country. Households in which the head (typically male) has attended school appear more likely to send their children to school as well.

Over 90% of households with children between the ages of 7 and 15 appear to own a family business. The survey classifies households' businesses into two groups: Farm Enterprises (FE) and Non-Farm Enterprises [13]. Out-of-school children are at least twice as likely to belong to a family with a household business than their counterparts who are enrolled in school. More than half of families with UBE-aged children appear to have both businesses. In addition, it appears belonging to a family not owning a farm is highly correlated with school enrollment rate. 37% of children without family farms attend school full-time while only 7% of the children in this group do not attend school. This is consistent with the idea that farming (as is typically practiced in Nigeria) is more time intensive than typical non-farming employment activities.

Table 1.2

Features of Nigerian Children Ages 7-15 and their households (Sample Means and Proportions)

Features	Not Attending School	Working and Schooling	Schooling without Working
Age	10.8	11.6	10.7
Age of Household Head	50.9	51.3	50.2
Birth Position	3.0	2.5	2.8
Household Head Years of Schooling	2.1	5.3	7.5
Household Income (in Naira)	27891.1	54987.7	69797.6
Household Members Under Age 5	1.1	0.7	0.9
Household Size	8.9	7.8	8.0
Population Size	6352222.0	2445734.0	31872656.0
Years of Schooling	1.2	4.7	4.2
Rural	0.2	0.1	0.7
Urban	0.0	0.0	0.9
Female	0.2	0.1	0.8
Male	0.2	0.1	0.8
NC	0.2	0.1	0.7
NE	0.3	0.1	0.7
NW	0.3	0.0	0.7
SE	0.0	0.1	0.9
SS	0.0	0.1	0.9
SW	0.0	0.1	0.9
Non-Farm Enterprises (NFE)	0.0	0.0	0.9
FE & NFE	0.2	0.1	0.7
Farm Enterprises (FE)	0.2	0.1	0.7
None	0.1	0.1	0.8

Note:
Source: Nigeria General Household Survey Panel 2015/16 and author's calculations

The earnings regression uses a generalized linear model to estimate an earnings function (equation 3) below. We find that the log earnings model covariates largely have the appropriate signs. The main aim of the model is to estimate the effect of occupational status on child earnings, specifically for those children who work either in an FE or NFE. Children who are working and schooling earn less than children who work and don't attend school. It appears children who work while attending school earn about 22.7% less than out-of-school children who work. This variable appears to be statistically and economically significant which is important as this is used within the model estimation.

Given the prevalence of informal household businesses within the Nigerian economy, the earnings model is estimated while controlling for class of family business ran. We include a variable for type of household business with 4 categories: Non-Farm Household Enterprises (base case), Farm Household Enterprises, both (NFE & FE) and neither. As expected, children working in families with non-farm household enterprises appear to earn more. Children with larger households earn less. Children in states with higher median state household income and educational expenditure also earn more. There are differences in the 6 geo-political zones that match our expectations. For instance, the rates of work are significantly higher in the Northern than in the Southern geo-political zones (GPZ). This is expected because the school enrollment rates are markedly higher in the South than in the North. Using the North Central GPZ as the base case, the value of child labor appears to be higher and statistically significant in the North East (+ 0.42%) and in the North West (+ 1.363%) but is lower in the South East (-0.448%), South South (-0.133%) and South West (-0.770%).

Table 1.3

	Logged Monthly Earnings
	Log(Income)
Age	0.004
	(0.074)
Male	-0.021
	(0.081)
Urban	-0.100
	(0.169)
Birth Order	-0.001
	(0.031)
Household Members Under Age 5	0.069*
	(0.041)
Farm and Non-Farm Household Enterprises	-0.654***
	(0.180)
Farm Household Enterprises	-0.110
	(0.192)
No Household Enterprise	0.477*
	(0.273)
Household Size	-0.034*
	(0.019)
Years of Schooling	0.051
	(0.066)
Age minus Years of Schooling squared	0.001
	(0.004)
Working and Schooling	-0.227**
	(0.100)
Log(Median State Household Income)	0.596***
	(0.185)
Log(Median State Educational Expenditure)	0.606***
	(0.104)
North East	0.420**
	(0.167)
North West	1.363***
	(0.189)
South East	-0.448**
	(0.210)
South South	-0.133
	(0.191)
South West	-0.740***
	(0.227)
Constant	-1.930
	(1.901)
Observations	884
Log Likelihood	-1,387.472
Akaike Inf. Crit.	2,814.944
Note:	*p<0.1; **p<0.05;*** p<0.01

Log Earning Regression for Children Ages 7-15

The results of the occupational choice model, in large part, are in line with standard development economic theory with a few noteworthy and explainable departures. The average marginal effect estimates are presented in table 1.4. The model estimates occupational choice for each child based on the set of covariates in the table 1.3. The reference category is not attending school. All else equal, household income is linked with higher schooling with small effect sizes for children working and schooling and those schooling without working. Higher predicted earnings is associated with a lower probability of schooling without working. For (-0.000002) each 1 Naira (~ US 2.5 cents) increase in the child's predicted earnings the odds of switching going from being out of school to both schooling and working reduces by a factor of 0.000002.

Consequently, a 5000 Naira increase (~ USD 14) i.e. the standard transfer amount proposed by the government would be associated with a 0.12 factor reduction in the probability of the same shift. Increase in age is associated with more schooling. Boys are less likely than girls to be attend school full-time but more likely work and school as well as work full time than women. In a largely agrarian society, it may indicate that boys are over-selected for labor market work while women are perhaps more likely to do domestic work and attend school.

Children in families with household members under age 5 are on average less likely to be working and schooling (-0.010) than they are to be schooling without working. In addition, an increase in the number of years of schooling has a negative average marginal effect on the likelihood of attending school without working (-0.040). Consistent with the earnings model and controlling for other factors, children appear less likely to enroll in school in the Northern GPZs, East (-0.018) and West (-0.018) while in the South, East (0.109), South (0.074) and West (0.092) enrollment appears to be higher.

As previously explained, multinomial logit estimations only permit identifying differences in effects of dependent variable options (based on a reference case, Not Schooling in this case). However, simulating any cash transfer of this kind requires estimation of the marginal utility with respect to the reference case. Based on equations 15, we get the following results:

$$M = e^{m} = 0.77; D = 0.57$$

Consequently, for every 1 NGN contributed by a child working in the labor market without schooling, a child splitting time between school and work contributes 0.77 kobo (77% of 1 NGN) to their family. In other words, a family loses 23% of a child's labor productivity for choosing to send the child to school while participating in a family owned business. The D parameter explains that domestic labor productivity of children in school is only about 57% of their potential labor market earnings. This means that the marginal product of child labor for those in schooling and working is different from that of those in full-time work. This implies that the estimated value of nonmarket work by children studying full-time is about 20% less than the market value of working while enrolled in school. This may indicate that there is selection on observables into market work. M and D as well as $\hat{\alpha}$ and $\hat{\beta}$ are used in the structural estimation of the marginal utilities of the occupational choices with respect to household income and potential child earnings. The results of estimations are as follows:

$$\alpha_0 = 0.000081; \alpha_1 = 0.000081; \alpha_2 = 0.000078$$

$$\beta_0 = -0.000001; \beta_1 = -0.000003; \beta_2 \approx 0$$

The α's show the estimated effect of 1 NGN increase in household income (minus income of the child) on school choice. The β's show the estimated effect of 1 NGN increase in child's income on school choice. The effect sizes appear to be small but given the Nigerian currency is consistent with the academic literature as we show in the result section. The direction of the signs from the structural model parameters are consistent with economic theory. The marginal utility of each occupational choice with respect to income increases in the direction of more schooling. In addition, the marginal utility of each occupational choice with respect to the expected contribution of each child to income is declining in the direction of more schooling. The model therefore implies that each marginal increase in the expected earnings of the child increases the utility accruing to the family for not sending their child to school while at the same time receiving a disutility for not sending their child to school. The direction of these two results are essential within the simulation as the likelihood of a switch is solely dependent on the relative magnitude of the coefficients.

Table 1.4

Table 1.4: Occupational Choice Model for Children Age 7-15: Marginal Effects and p-Value Estimates

Variable	Classification	Marginal Effect	p-value
Household Income	Working and Schooling	0.0000000	0.534
	Schooling without Working	0.0000005	0.000
Predicted Child Earnings	Working and Schooling	0.0000011	0.174
	Schooling without Working	-0.0000024	0.043
Age	Working and Schooling	0.0008206	0.879
	Schooling without Working	0.0652397	0.000
Male	Working and Schooling	0.0249000	0.000
	Schooling without Working	-0.0091767	0.309
Urban	Working and Schooling	-0.0402172	0.000
	Schooling without Working	0.1141566	0.000
Birth Order	Working and Schooling	-0.0046129	0.067
	Schooling without Working	0.0029878	0.357
Age of Household Head	Working and Schooling	-0.0000885	0.750
	Schooling without Working	-0.0002413	0.563
Household Members Under Age 5	Working and Schooling	-0.0103273	0.010
	Schooling without Working	0.0166703	0.002
Household Size	Working and Schooling	-0.0017020	0.260
	Schooling without Working	0.0004956	0.813
Years of Schooling	Working and Schooling	0.0093217	0.072
	Schooling without Working	-0.0397052	0.000
Log(Median State Household Income)	Working and Schooling	0.0359868	0.003
	Schooling without Working	-0.0783961	0.000
Log(Median State Educational Expenditure)	Working and Schooling	-0.0078946	0.060
	Schooling without Working	0.0596531	0.000
North East	Working and Schooling	-0.0299072	0.007
	Schooling without Working	-0.0179503	0.305
North West	Working and Schooling	-0.0524019	0.000
	Schooling without Working	-0.0182548	0.284
South East	Working and Schooling	0.0044292	0.793
	Schooling without Working	0.1086415	0.000
South South	Working and Schooling	0.0107867	0.531
	Schooling without Working	0.0742552	0.000
South West	Working and Schooling	0.1927988	0.000
	Schooling without Working	0.0919207	0.024

[a] Note: Pseudo R-squared = 0.344; N = 6,508
[b] Source: Nigeria General Household Survey Panel 2015/16 and author's calculations

1.5 Estimation Strategy: Ex-Ante Simulations of Conditional and Unconditional Cash Transfer Programs in Nigeria

CCTs and UCTs and other social safety net programs and policies aim to reduce poverty as well as other goals by means of targeted transfers. Ex-ante analysis, under strong assumptions as we employ in this model, are one method of measuring the likely effects of future program designs on school enrollment. It also allows policymakers to weigh risks and benefits of differing program specifications particularly under restrained circumstances.

In the Nigerian case study, the federal government had originally planned a conditional cash transfer program. One such program would have provided funds to eligible families for children attending school between the ages 7-15. Due to difficulties monitoring compliance, however, the government opted for unconditional transfers. The question then becomes how much less effective, in theory, these programs (specifically with regards to schooling outcomes) are likely to be if all eligible families receive transfers irrespective of school attendance. How much less frequently are families expected to switch occupational status of their wards under the unconditional case assuming the transfer amount remains the same as under the conditional regime? This is important for strategizing towards meeting the 2030 Global Sustainable Development Goals and the 2063 African Sustainable Development Agenda Goals which both have a strong poverty focus. This is important for strategizing towards meeting the 2030 Global Sustainable Development Goals and the 2063 African Sustainable Development Agenda Goals which both have a strong poverty focus. We apply the model approach, behavioral parameter values $(\alpha, \beta, \gamma, M, D)$ and decision rules (equations 9-12) described previously to the Nigeria's General Household Survey 2015/2016 population. Under this approach, we can simulate the counterfactual distributions of occupational status generated by running both conditional and unconditional cash transfer programs. We begin by running a simplified version of the intended cash transfer program (results table 1.5). In this program, each state government would be providing 5000 NGN to the bottom 10% poorest households on the condition that they sent their child to school (CCT program). The unconditional comparison would simply be a program in which the same bottom 10% of households receive a transfer regardless of child school enrollment (UCT program). We present the results of both programs below.

We test 5 other counterfactual designs varying transfer amount, program coverage, and transfer amount per child. In Design 1, the government offers 5000 NGN to the bottom 10% poorest households in Nigeria (UCT program). In the second scenario, the same transfer amount is offered the poor households who send their children to school (CCT program). Design 2 is similar to design 1, however, the transfers in both CCT and UCT programs are provided on a per child basis. Design 3 is the same as design 1 except the bottom 20% of households are eligible for both UCT and CCT programs. Design 4 provides double the transfer amount per child to the bottom 10% of households being eligible for both transfer types. Design 5 is like design 4 except that the program now reaches the bottom 20% rather than the bottom 10% of households. In the subsequent section, we present the results of the counterfactual simulations.

Design Number	Transfer Amount (in NGN)	Per Household (HH)/Per Child	Coverage Rate
1	5000	Per household	10%
2	5000	Per child	10%
3	5000	Per household	20%
4	10000	Per child	10%
5	10000	Per child	20%

Note: In each design, a UCT will provide blanket transfers meeting the features described. A CCT will provide the transfers on the condition of school attendance for a child (designs 2, 4,5) or at least one child of the household attends school (designs 1, 3).

Table 1.5

Table 1.5: Simulated Household Income Effect on Occupational Choice Ages 7-15

	Schooling and Working	Not Attending School	Schooling without Working
Unconditional Programs			
Schooling and Working	100.00	0.00	0.00
Not Attending School	0.00	97.97	2.03
Schooling without Working	0.00	0.00	100.00
Conditional Programs			
Schooling and Working	100.00	0.00	0.00
Not Attending School	10.37	30.39	59.24
Schooling without Working	0.00	0.00	100.00

[a] Source: Nigeria General Household Survey Panel 2015/16 and author's work

Table 1.6

Table 1.6: Alternative Unconditional and Conditional Cash Transfer Program: Simulated Effects on Occupational Status of Children Ages 7-15 (percent)

	Baseline	Design 1	Design 2	Design 3	Design 4	Design 5
Conditional Programs						
Schooling and Working	5.4	5.6	5.8	5.9	6.4	8.0
Not Attending Schooling	23.1	21.6	21.6	20.0	21.0	18.3
Schooling without Working	71.5	72.8	72.6	74.1	72.6	73.7
Unconditional Programs						
Schooling and Working	5.4	5.4	5.4	5.4	5.4	5.4
Not Attending Schooling	23.1	23.1	23.1	23.1	23.1	23.0
Schooling without Working	71.5	71.5	71.5	71.5	71.6	71.6

Table 1.7

Table 1.7: Intensity of Poverty amongst Poor Households

Scenario	Design 1	Design 2	Design 3	Design 4	Design 5
Baseline	0.4014502	0.4014502	0.4014502	0.4014502	0.4014502
UCT	0.3675850	0.3675850	0.3354282	0.3337198	0.2694062
CCT	0.3776087	0.3774998	0.3570265	0.3535495	0.3123561

Table 1.8

Design Number	UCT Cost in million USD (% 2017 GDP)	CCT Cost in million USD (% 2017 GDP)	UCT eligible households (in millions)	UCT-CCT Cost Ratio
1	7.54 (0.0020%)	5.91 (0.0016%)	20.8	1.275
2	14.4 (0.0038%)	12.9 (0.0034%)	20.8	1.110
3	15.3 (0.0040%)	12.3 (0.0032%)	40.6	1.250
4	28.7 (0.0076%)	28.3 (0.0075%)	20.8	1.013
5	61.6 (0.0163%)	60.6 (0.0161)	40.6	1.018

Simulation Results

From the results in table 1.5, around 6-7 out of 10 children who are not attending school before a conditional program will switch into attending school full-time after the program. In the unconditional case, the chances are significantly less with less than 3% of the children who receive UCTs attend school due to the transfer. However, school enrollment appears to be far less responsive in the UCT case than in the CCT cases.

One theory is that perhaps the transfer amount is too low and agents within the simulation will be more responsive to larger transfer amounts. Table 1.6 presents results of alternative simulations comparing the effects on school enrollment of CCT and UCT programs. The baseline column shows the simulated state of occupational choice for children between the ages of 7-15 in Nigeria.

In terms of school enrollment, the UCT programs do not appear to change educational enrollment amongst the poorest 10% of households across all the states of Nigeria. However, these results will understate how effective unconditional programs are in increasing educational enrollment since in practice program staff tend to encourage families to use the cash inflow to enroll their children in school and UCT programs often have “soft conditions” which are uncaptured in the modelling. We do not know the extent to which this is effective, and our modelling does not take this into account. From table 1.6, the CCT program designs appear more effective at increasing school enrollment when the government increases coverage (comparing designs 1 and 3) than when government increases the transfer amount per child (comparing designs 1 and 2). The combination of both increases in transfers per child and coverage appear to lead to larger increases as seen in design 3 (table 1.6). In the final two designs, the transfer amount is doubled, per household in design 4 and per child in design 5. In both programs, we see changes that suggest that increasing coverage is yet again more effective than increasing the transfer. In design 4, doubling the transfer amount only increases the rate of school enrollment by 0.6% (over design 1) while in design 5, doubling the coverage rate increases enrollment by 3.3% (over design 2).

In table 1.7, we measure the intensity of poverty amongst poor households. We compute this by taking the absolute difference between household daily income and the international poverty line ($1.90 per day per person) as a fraction of the poverty line. As expected, we find that UCTs are more effective at reducing poverty amongst the extreme poor than are the CCT programs since their coverage is less. Coverage appears to drive larger reductions in the intensity of poverty than does simply providing transfers per child (comparing Designs 3 and 1).

In terms of programs costs (table 1.8), if the government prefers to run a CCT and hence absorb the cost of monitoring, it should be willing to absorb 27.5%, 11%, 25% 1.3% and 1.8% of the corresponding UCT program in designs 1-5 respectively. The per child CCT programs appear to lead to higher school enrollment leading. Consequently, gross per child transfer program costs are higher and as such there is a smaller gap between UCT and CCT costs which explains the lower ratios seen in designs 2, 4 and 5.

This study is not without certain limitations. Estimating the effects of changes in income on education as well as poverty is a complex issue and as such our results are meant to stand as ballpark estimates with which we simulate likely outcomes of policies. The decision to attend school is often a complex intertemporal iterative decision-making process which we have for purposes of simplicity reduced to a static unitary decision within the household. The study introduces utopian style cash transfer programs. CCTs usually involve imperfect monitoring and significant non-compliance while UCTs often carry soft conditions and cash is not as freely provided as described in this study.

Conclusions

The Federal Government of Nigeria in its bid to reduce poverty and address other socio-economic issues such as low educational enrollment originally planned to implement conditional cash transfer programs. Due to its inability to monitor program compliance, it opted for unconditional benefits that would only provide transfers based on income eligibility. In this study, we sought to measure the likely effects of UCT and CCT program scenarios on educational enrollment and the intensity of poverty amongst the poor. We found that CCT programs are far more effective at increasing enrollment in primary education than UCTs, but the UCTs are more effective at poverty reduction. We estimate the costs of both program types (UCT, CCT) under the multiple program designs which could help a government decide whether to invest in monitoring the CCT program rather than taking on a UCT program that appears to provide less improvements to educational participation for children.

There are a few key results worth noting from this study. Increasing program coverage and providing transfers per child in school appear to increase educational enrollment more than do increasing the transfer amount. However, increasing program coverage almost doubles transfer costs over providing transfer per child all other features equal. Proposing to provide transfers for each child that attends school appears to increase the risk of working while attending school. While this increases the overall rates of schooling, it does not necessarily improve the rates of child labor and this is important depending on the priorities of policymakers. We show that investing between 1.3% – 27.5% (depending on the program design) of transfer costs in an education based CCT program maybe more beneficial to a Nigerian administration with priorities to increase basic education participation amongst the poor.

Chapter 2: Predicting the Demand for Dementia Care: Simulating Dementia Incidence in the United States

Ifeanyi Edochie, Carolyn M. Rutter PhD, Esther Friedman PhD, Mary Slaughter PhD

RAND Corporation

December 19th, 2019

2.0 Introduction

By 2035, adults 65 and older are expected to outnumber children less than age 18 for the first time in US history [14]. As the U.S population ages, the overall prevalence of diseases associated with older age will increase. Alzheimer's disease is currently the sixth leading cause of death in the United States and the fifth leading cause of death among those over age 65 [15]. Alzheimer's disease involves progressive cognitive decline that limits independent functioning [16]. At some point in the disease progression, most people with the disease require assistance with personal care - help with everyday activities, also called "activities of daily living." These activities include bathing, dressing, grooming, using the toilet, eating, and moving around -- for example, getting out of bed and into a chair. The majority of care for these types of needs of older adults with dementia is provided by family and friends [17]. While Medicaid provides support for personal care and other long-term care needs, only people with the lowest level of assets qualify for Medicaid. The lack of formal care access to most of the U.S. population may explain why much of Long Term Services and Supports (LTSS) is either paid out of pocket or shouldered by family [18]. Currently, more than 15 million Americans provide care to family members or friends with dementia [19].

Recent demographic changes could affect the likelihood that people with Alzheimer's disease who need care will have the family members available to provide it. Declines in U.S. fertility have persisted since the 1950s, though the rate of change has slowed. In 2018, the U.S fertility rate was about 1.7 children born per mother.[20] At the same time, medical innovations have led to increased lifespans. While the population of children under age 18 continues to decline, older adults are living longer, healthier lives which could suggest that spouses may be able to assume caregiving roles longer. Changes in marriage and remarriage could also alter the availability of kin to provide care.

In our broader work on this topic, we sought to understand how demographic factors – fertility, marriage, and mortality – contribute to the family care that will be available to older adults with Alzheimer's disease in the future. To do this requires linked information on two projected trends: (1) family availability (caregiving supply) and (2) dementia incidence (caregiving demand).

Simulation methods provide a tool for predicting both these trends. Microsimulation models simulate disease trajectories for individual agents by applying probabilistic rules governing state transitions [21]. Models can be used to simulate future transitions in an existing population [22-24], or can be used to simulate future transitions for a hypothetical population[21], for example, a population that represents the U.S. population in 2060. To simulate transitions, microsimulation models incorporate information from published research and expert opinion. Microsimulation models assist policy makers by making predictions based on synthesis of high-quality evidence across multiple sources. Models are especially useful when they are used to generate predictions under different assumptions, which can include assumptions about the etiology of disease or policy interventions.

In this paper, we develop a microsimulation model to predict the future demand for dementia care. To project the supply of family care, we start with an established socio-demographic model for simulating a synthetic population with familial networks [25]. We then introduce a dementia module to capture caregiving demand, which simulates the onset of and death from dementia for agents in the synthetic population. In Section 2.1 we describe the Socsim kinship model. In Section 2.2 we provide a detailed description of the dementia module, including the model parameters, methods used to calibrate model parameters, and our approach to model validation. In Section 3.0 we provide results, including the parameters used in the dementia module and validation of the combined Socsim-D (Socsim+ dementia) model. We conclude with a discussion of how the Socsim-D model can be used to inform health policy. While our focus is on linking our new dementia model to Socsim, in the discussion we describe other possible applications of this model to other populations.

2.1 Methods

To project the demand for care requires the linking of two models: a model of the projected family caregiving supply as well as caregiving demand (dementia). To achieve this, we developed a dementia module that can be used in conjunction with an existing simulation model of kinship, Socsim, to explore the joint trends in dementia prevalence and the availability of family caregivers.

2.1.1 The Socsim Model

SOCio-demographic MicroSIMulation, typically referred to as Socsim, is a continuous closed-population kinship simulator developed at University of California at Berkeley's Demography Lab in the 1970s [26]. The Socsim model simulates a synthetic population based on an initial population and demographic rates. Every month, each individual in the simulated population faces the risk of a number of events occurring, including childbirth, death, and marriage. Other constraints may also be included in the simulation program (e.g. minimum interval of time between births etc.). To build our population, we started with a sample of two hundred thousand white and black individuals from the U.S. 1880 Census. Our sample included age, race and marital status. We used yearly mortality, fertility, and marriage rates from 1880 to 2060. The final Socsim population that is returned includes a list of every family member, their sex, date of birth, and the years they lived. In addition, family networks (siblings, spouse and children across generations) are simulated onto the Socsim population. The model deals with divorce by assigning divorcees a special code in a marriage file. It also allows for remarriage in which case divorcees have another record in the marriage file with a later date when their marriage began and ended.

We used the Socsim model to simulate a population of approximately 7 million that is representative of the U.S population between 1910 and 2060. To project the need for and availability of family caregivers, we simulate agents born in the simulation time period mentioned. The Socsim model assigns each agent a race (black or white), sex, dates of birth and death, and kinship networks. The ability to realistically simulate kinship networks allows us to estimate the potential supply of family care for dementia patients once demand estimates are computed. Fertility and mortality rates used in the Socsim models are calibrated so that the synthetic population matches actual US Census data for the years 1915 to 2015. We also match US demographic projections in the same parameters between 2015 and 2060. We use the model to project the US population through 2060 by assuming state-steady-state assumptions.

2.1.2 Projecting Dementia to the Socsim population.

We developed a dementia module which simulates the onset of and death from dementia for agents in the synthetic population in a way that is consistent with the epidemiology of dementia but does not reassign death dates. We do not reassign death dates to maintain the validity of the Socsim model calibration.

There are four components of the dementia module: 1) simulation of lifetime prevalence of dementia for each birth-cohort; 2) simulation of the age of onset for agents with a dementia diagnosis, based on sex and race; 3) simulation of the age of death with dementia for agents with a dementia diagnosis, based on age at diagnosis, sex, and race; and 4) assignment of dementia age at onset to agent in the synthetic population, based on birth cohort, sex, race, and age at death.

2.1.3 Lifetime Prevalence of Dementia

Our first step is to determine how many Socsim agents in each birth cohort will develop dementia at some point in their lives, that is, their lifetime risk of dementia. This simulation needs to account for the dependence between the lifetime risk of dementia and life expectancy because individuals who live longer have more opportunity to develop dementia. For example, published estimates of the lifetime risk of dementia find that women have a higher lifetime risk (25.2%) than men (18.7%) [27].

We assume that the lifetime incidence (LI_{sb}) of dementia among agents that are alive at age 65 is proportional to their life expectancy at birth (LI_{srb}), where s indicates sex (male or female), r indicates race (black or white), and b indicates birth-year cohort: $LI_{sb} = \alpha_s LE_{srb}$. We allow the constant of proportionality, α_s, to differ for men and women.

To estimate the model parameter α_s, we combine sex-specific estimated lifetime dementia incidence in 2014 from [27] with sex-specific estimated life expectancy for agents that are 65 in 2014, corresponding to birth-year $b = 1949$: $\alpha_s = \frac{LI_{s,1949}}{LE_{s,1949}}$. We then use this estimate of α_s to calculate the lifetime incidence for birth cohorts from $b = 1910$ to 2015 using $LI_{sb} = \alpha_s LE_{srb}$. Here, we assume that dementia risk, as captured by α_s, is unchanged for birth cohorts from 1910 through 2014 (the last cohort to turn 65 before 2060), and that at a population level black/white differences in lifetime dementia risk are attributable to differences in life expectancy.

In summary, in each simulated year the number of Socsim agents who are 65 that will have a diagnosis of dementia before death is based on applying α_s to life expectancy calculated using the synthetic population. We assume α_s is fixed across years, and so any changes in lifetime incidence of dementia are driven by changes in life expectancy.

2.1.4 Age at Dementia Onset

We simulate age at dementia onset for each age, race, and birth cohort using an estimated cumulative incidence of dementia onset at age a, among individual who are diagnosed with dementia in their lifetime. $I_{srb}(a)$. Birth cohort, *b*, is stratified into eight 20-year groups, from 1915 to 2060 resulting in 32 strata. Cumulative incidence describes the rate at which new dementia cases arise in the population, and depends on three other factors: the probability of surviving up to age a, $S_{srb}(a)$; the probability of remaining free from dementia up to age *a,* $p_{sr}(a)$; and the annual age-specific dementia incidence rate, $i(a)$. Combined, $S_{srb}(a)$ and $p_{srj}(a)$ provide the probability that an agent is at risk for developing incident dementia.

To simulate age at onset, we start with the synthetic population to estimate $S_{srb}(a)$ for age $a \in [65,100]$ in each race-sex-birth-cohort stratum. In other words, given our synthetic population, $S_{srb}(a)$ is known. We estimated the probability of not having a dementia diagnosis at age a using a linear regression model applied to data from the RAND Health and Retirement Survey [28]. [29] predict the likelihood of dementia among Aging, Demographics and Memory Study (ADAMS) sample respondents using an ordered probit regression towards projecting the costs of dementia care. The outcome variable of their model takes three values: 1 if dementia positive, 2 if respondent possesses cognitive impairment without dementia (CIND) and 3 if the individual possesses normal cognition. The outcome variable is assumed to be dependent on demographics as well as measures of Activity of Daily Living and Instrumental Activities of Daily Living (IADL) and cognitive functioning measured in previous HRS waves. Of the three probabilities in their analysis, we estimate the probability of being demented. $p_{sr}(a)$ is simply the complement i.e. the probability of being dementia free at a given age. We use a quasi-binomial regression specification dependent on age, race and sex of the following form:

$$1 - p_{sr}(a) = \beta_0 + \beta_1 a + \beta_2 a^2 + \beta_3(r = 1) + \beta_4(s = 1)$$

We treat this function as a model input rather than a calibrated parameter. That is, we assume that the model for $p_{srb}(a)$ is fixed and known given $\hat{\beta}$. The estimation yields, $\hat{\beta}_0 = 2.945,\ \hat{\beta}_1 = 0.382,\ \hat{\beta}_2 = -0.002,\ \hat{\beta}_3 = -1.08,\ \hat{\beta}_4 = 0.021$]. The direction of the coefficients appears to be consistent with the literature. Age is positively correlated with the likelihood of a dementia diagnosis while the rate of increase is decreasing with age. Whites have a lower prevalence of dementia and older men appear to have a higher dementia risk than their female counterparts. Therefore, we only treat the vector of annual age-specific dementia incidence rates, $i(a)$, as a set of calibrated parameters.

For each sex, race and birth cohort, the strata-specific cumulative incidence function, $I_{srb}(a)$ is calculated by computing the product $i(a)p(a)_{sr}S_{srb}(a)$, taking its cumulative sum and then rescaling the values obtained within each stratum such that every value is transformed to a number between 0 and 1, i.e. the rescaled value for cumulative dementia incidence at age a is scaled by dividing by $I_{srb}(100) = i(100)p_{sr}(100)S_{srb}(100)$. Then, for agents that receive a lifetime diagnosis of dementia, we simulate age at onset (A) using an inverse cumulative density function (CDF) lookup i.e . we assign each case of dementia a uniformly distributed variable, $r \in U(0,1), A : r = I_{srb}(a)$. The age of onset is defined by the age in a specific stratum corresponding to cumulative dementia incidence that equates to an assigned uniformly distributed random variable for a specific case of dementia.

In summary, age at dementia onset (A) depends on sex, race and birth cohort and is based on the cumulative incidence function, $I_{srb}(a)$, which in turn is a function of annual age-specific dementia incidence rates $i(a)$ which are treated as unknown model parameters that are calibrated.

2.1.5 Age at Death with Dementia

In 2014, Alzheimer's was the sixth leading cause of death [30]. The increased risk of death after diagnosis with dementia relative to individuals not diagnosed with dementia is estimated by the log-hazard ratio, λ, which ranges from 1.4 and 3.2 within the dementia literature [31-36]. The literature also shows that there are age and race dependencies with regards to survival by dementia status. Therefore, our model introduces race and age-at-diagnosis (A) specific log-hazard ratios, λ_{Ar}, obtained from [15].To simulate survival time after a dementia diagnosis, we combine age and race specific log-hazard ratios estimated in studies of individuals with and without a dementia diagnosis[15] with cumulative hazard functions that are estimated from the synthetic population. We estimate the cumulative hazard function separately for each race, sex and birth-cohort using the Nelson-Aalen estimator (NAE) [37]:

$$\lambda_{Ar} = \sum_{t_a < t} \frac{y_a}{d_a}$$

where y_a is the number of agents in the race-sex-birth-cohort strata that are alive at age a (the risk set) and d_a are the number in the risk set that die before their $(a+1)^{st}$ birthday. The corresponding all-cause cumulative survival probability, i.e. the probability of surviving past age a given the agent is alive at age a, can be expressed as:

$$S_{srb}(a) = e^{-\Lambda_{rb}(a)}$$

where $\Lambda(a)$ is the empirical hazard and is based on survival in the synthetic population. Expressing the survival in this format allows us to estimate the dementia-specific cumulative survival probability, $S(a|D=1)$, using a proportional hazards model:

$$S_{srb}(a|D=1) = e^{-\lambda_{Ar}\Lambda_{rb}(a)}$$

where λ_{Ar} is the log-hazard ratio associated with survival after a dementia diagnosis. We assume that the log-hazard ratio varies by race and by age at dementia onset with 6 age groups (64-69 years; 70-74 years; 75-79 years; 80-84 years; 85-89 years; and 90 years and older), resulting in 12 different log-hazard ratios. Next, for each sex, race, and birth cohort we simulate M_{srb} ages of dementia onset using and the cumulative incidence function. Finally, we assign M_{srb} ages of death with dementia.

In summary, to simulate age at death from dementia, we use the synthetic population to estimate the empirical cumulative hazard, $\Lambda_{rb}(a)$, for each race and birth-cohort. We then apply a proportional hazards model to this cumulative hazard, using log-hazard ratios that vary by age at diagnosis and race, λ_{Ar}. These log-hazard ratios are calibrated.

2.1.6 Simulating Agents with Dementia

Once we have a set of dementia characteristics, we match these within race-sex-birth-cohort strata to agents in the synthetic population, based on their integer ages at death.

2.2 Model Calibration

The newly constructed Socsim Dementia model (Socsim-D) contains three unknown parameter sets that we calibrate: α_s, the ratio of lifetime dementia incidence to life expectancy among agents that survive to age 65 (two values, for male ($s = 1$) and female ($s = 0$)); $i(a)$, the annual age-specific prevalence (7 values for age groups 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, and 95-99); and λ_{Ar}, log-hazard ratios that drive the increased risk of death after dementia diagnosis (12 values for black and white non-Hispanic agents and for 6 age at diagnosis groups: 65-69, 70-74, 75-79, 80-84, 85-89, 90+). We start by setting these unknown parameters to published estimates. These baseline parameters values were obtained from lifetime incidence rates of the 1949 birth cohort and age-specific annual incidence estimated[27] and age-and race specific log-hazard ratios published[15].

Model calibration is based on selecting model parameters that result in Socsim-D model predictions that are consistent with calibration targets. We updated our baseline parameter values using a simple search, or 'hand tuning'. Calibration was guided by deviance statistics used to measure the distance between targets and model predictions. We updated each of the three parameter blocks, rather than updating each of the 21 parameters individually.

2.2.1 Calibration Targets

We calibrated our model based on the ability of the Socsim-D model to predict age specific dementia prevalence and survival time after dementia diagnosis. We surveyed many studies that provide estimates of dementia prevalence and survival. We found five studies published in the last ten years that reported age-specific dementia prevalence for three age-groups: 65-74,75-84, and 85 and older [38-42]. We selected these studies (having surveyed several) because they all had large sample sizes and used well-established methods to determine participants' dementia status specifically for the US. In addition, the three age groupings settled upon appear to be the most standard in the field and we only selected studies that follow this exact age grouping to allow for easier comparisons of studies. Figure A1 shows dementia prevalence by age across these studies. This ranges from 2.8 to 4% in age group 65-74, 10 to 17.6% for the 75-84 years and 24.9 to 47.5% for the oldest. Due to the wide variation in the prevalence estimates, particularly for the 85+ group, we chose to focus on the study[41] as our primary calibration target: its prevalence falls in the middle of the range of estimates, it is a large and representative sample, based on a 20% sample of the Medicare population in 2008. The sample size of [41] in the age group for those 85 years and over is important because it guarantees a more consistent estimate for dementia prevalence.

We use a recent study by [15] to inform both our log-hazard ratios, stratified by race- and gender, and to provide calibration targets for survival time after diagnosis with dementia. This study used Kaiser Permanente Northern California (KPNC) data and included 18,778 African American patients and 20,649 Caucasian patients over the age of 64 who were followed from January 1st, 2000 to December 31st, 2013 for dementia diagnosis and subsequent survival. The median survival time after diagnosis was 3.6 years for African Americans (1.1-7.6) and 3.1 years for Caucasians (0.9-.6.3). We also used the [15] paper to compare Socsim-D model predictions to estimated hazard ratios from [15], for people with compared to those without dementia.

2.3 Results

2.3.1 Calibration and Model Results

Model calibration began with three runs, corresponding to three different assumptions for the annual dementia incidence rate. Two of the annual incidence rates are based on multistate model estimates and a systematic review published by [43]. A third annual incidence rates set is based on an analysis of dementia incidence using the RAND HRS data [44]. Model predictions from all three of these runs resulted in simulated age-specific dementia prevalence that was too low at all ages, and survival times that were far lower than targets (See Table A1), which is why calibration was necessary.

Step 1: The goal at the first calibration step was to increase prevalence at younger ages by increasing α_s, and to also increase prevalence at older ages by increasing survival time after diagnosis, by decreasing log-hazard ratios, correspond to dementia having a smaller effect when applied to all-cause survival. We reduced the magnitude of λ_{Ar} by assigning it to the lower 95% confidence limits provided by [15], We anticipated reducing these log-hazard ratios because they were estimated based on comparing survival for those with and without dementia but were applied to all-cause survival. We also decreased α_s, calculating it using the lower 95% confidence limits for lifetime dementia incidence, $LI_{s,1949}$, provided by [27].

After this change, dementia prevalence in the youngest age groups was improved, but prevalence in the oldest age groups was still much too low. Simulated survival time was also approximately half the target values in whites and blacks (1.8 years and 1.7 years respectively). By the end of Step 1, we dropped the annual incidence rates in [44] as a potential candidate for $i(a)$ because it resulted in prevalence rates that were too high in the youngest age group and too low in the oldest age group.

Step 2: To further improve model fit to targets, we increased α_s, calculating it using the lifetime dementia incidence, $LI_{s,1949}$, that was halfway between the mean and upper bound 95% confidence limit.   We also reduced λ_{Ar} by one half.

This resulted in model predictions that were a good match to prevalence and survival time targets. The deviance between estimates in dementia prevalence and survival and their respect targets were lower as seen in Appendix Table A1. Appendix Table A1 shows all the results of all the steps of calibration beginning with the baseline results from the literature. The second column shows the result of each scenario run in terms of dementia prevalence. The following column computes a simple deviance estimate which is the squared difference between simulated dementia prevalence and calibration targets for summed across all the ages. The final two columns show the results of simulated survival time for each run as well as deviance between simulated survival times and target survival times. We selected the setting that uses the systematic review's annual incidence rate from [27] because they resulted in a better match to dementia prevalence targets.

Final model parameters are shown in Table 1. In this table, we show final calibrated values for each parameter with each of the sub-groups. These are the parameter values that currently drive the demand for dementia care estimated within the Socsim-D model.

2.3.2 Comparison of Model Predictions to Other Published Results

As shown in Figure 1, the Socsim-D model resulted in age-specific incidence rates that were consistent with published results. Socsim-D predicted log-hazard ratios for survival after dementia diagnosis relative to a dementia free population. This estimate is between 1.2 – 4, as is consistent with the dementia literature [31-36].

We compare Socsim-D's effective λ_{Ar} with age-race specific log-hazard ratios from [15] as well as the $\lambda's$ from the studies listed above. [15] estimates appear to be demonstrably too high as the resulting survival times due to a dementia diagnosis for Socsim agents were too short. However, Socsim-D's effective λ_{Ar} rates (in Figure 2) appears to be consistent with other work [31-36]. In addition, our hazard ratios must be lowered since as previously stated the literature estimates are survival ratios based on dementia status while Socsim's hazard ratios compare all-cause death (including dementia patients) with death after a dementia diagnosis. Effective hazard ratio estimates can also be found in Appendix Table A2.

2.3.3 Model Predictions of Future Dementia Prevalence

The Socsim-D model prevalence rates are similar to our calibration targets from [41] (Figure 3). We simulated both age-specific and overall population dementia prevalence from 2000 to 2060 (Figure 3). The Socsim-D model predicted a sharp decline in overall dementia prevalence from 2008 to approximately 2026. After about 2026, the model predicted a relatively rapid rise in dementia prevalence. We see from Figure 5 that the aging population is driving these increases in prevalence. The average age of elderly population (65 and over) declines steadily from about 73.7 years to 72.9 years between 2008 and 2020 in the same time dementia prevalence falls about half a percentage point. In the next 27 years, dementia prevalence consistently rises from about 5.7% to 7.1% while the average age of the same population monotonically increases from about 73 to 76 years. By 2047, dementia prevalence appears to fall slightly about 0.3% while the average of the population also drops to about 75.6 by 2057.

2.4 Discussion

We developed a module to simulate dementia incidence that we used in conjunction with Socsim, an existing intergenerational microsimulation model. The resulting composite model, Socsim-D, provides a tool for exploring trends in dementia prevalence, mortality and the availability of family caregiving. Our model results suggest changes in dementia prevalence in the period between 2000 and 2060 that are solely driven by the age dynamics of the population over age 65. In the periods of rising dementia prevalence, the average age of those over the age of 65 appears to rise and during periods of decline, the average age falls likewise. This is consistent with the fact that dementia prevalence is an increasing function of age.

We find that dementia prevalence appears to be stable between 2000 and 2006 and then begins to decline after 2008. Several studies report a reduction in dementia prevalence between 2000 and 2012 [40, 42]. Our results are consistent with these findings. [42] attributes the fall in dementia prevalence to "increased educational attainment and better control of cardiovascular risk factors". [40] find that demographic factors such as age might play a role as well. However, the magnitude of the contribution of these factors is unknown. Our results suggest that the age distribution in the population is and will continue to be an important factor driving dementia prevalence. Our model results predict a decline in dementia prevalence that will persist till about 2026. The model predicts a rise in prevalence rates until about 2047. This is driven by a consistent rise in the average of those alive and older than 65 within Socsim for the same time period. After 2047, the prevalence rate slightly declines as the elderly population becomes a little younger.

Our model is useful for a number of reasons. Socio-demographic microsimulation models can be useful here because they allow the application of scientific dementia information to be systematically applied to a synthetic population that has already been assigned continuous dates of birth and death, Socsim. The Socsim-D model is a new tool for assessing the potential effects of policies on the demand for care as well as the supply of family caregivers in the United States.

Simulating dementia incidence is not without its challenges. The key challenge is to account for age at death that is consistent with shorter survival for agents with dementia. Failing to account for this would result in a model that overstates the need of care. We apply a Nelson Aalen Estimator approach to Socsim survival probabilities such that age at death is consistent with a distribution of survival after a dementia diagnosis. In addition, we faced the methodological obstacle of assigning disease status without changing the survival distribution present in Socsim. This is important because in changing the survival distribution, the population ceases to represent the mortality in the United States. We dealt with this by separately assigning ages at dementia onset and death within a lifetime incidence set and using a simple search algorithm to find Socsim agents who best fit the characteristics of the lifetime incidences simulated.

Our analysis has some limitations. Ideally, we would have calibrated our model using a directed search algorithm or Latin hypercube sampling of the parameter space. Due to the computational limitations, we use an ad-hoc calibration method which nonetheless resulted in model predictions that were near our targets. In addition, our model for dementia risk was simple, depending only on age, race, and gender, but did not incorporate other factors that may be related such as educational attainment [42, 45] and there are other risk factors. In addition, the availability of new medical biomarkers[42] may also result in changes in the likelihood of a diagnosis over time. The difficulties in the disease diagnosis results in inaccurate observation of dementia onset with observational studies as diagnosis often occurs long after symptoms become noticeable. As a result, study estimates are likely to systematically underestimate disease incidence. In our use of published results from observational studies, Socsim-D does not address this problem.

Because our relatively simple model for dementia can be paired with a complex kinship simulator, it provides a useful tool for predicting the availability of family caregivers for patients with dementia in light of changing patterns of fertility and mortality. Future research will use the Socsim-D model to better understand both the future need for and availability of dementia care.

Tables and Figures

Table 2.1: Socsim-D, Final Calibrated Values for Dementia Module Parameters

Parameter	Variable Name	Subgroup	Final Calibrated Value	
Ratio of lifetime incidence to life expectancy	α_0	women	0.379	
	α_1	men	0.294	
Annual dementia incidence	$i(65-69)$	Ages 65 to 69	0.0025	
	$i(70-74)$	Ages 70 to 74	0.0042	
	$i(75-79)$	Ages 75 to 79	0.0079	
	$i(80-84)$	Ages 80 to 84	0.0148	
	$i(85-89)$	Ages 85 to 89	0.028	
	$i(90-94)$	Ages 90 to 94	0.0528	
	$i(95-99)$	Ages 95 to 99	0.0996	
Age to death after dementia diagnosis	λ_{black}(65-69)	Among blacks, from ages 65 to 69 years	3.1	
	λ_{black}(70 - 74)	Among blacks, from ages 70 to 74 years	2.755	
	λ_{black}(75 - 79)	Among blacks, from ages 75 to 79 years	2.365	
	λ_{black}(80 - 84)	Among blacks, from ages 80 to 84 years	2.17	
	λ_{black}(85 - 89)	Among blacks, from ages 85 to 89 years	1.915	
	λ_{black}(90+)	Among blacks, from ages 90+ years	1.5	
	λ_{white}(65 - 69)	Among whites, from ages 65 to 69 years	3.725	
	λ_{white}(70 - 74)	Among whites, from ages 70 to 74 years	3.145	
	λ_{white}(75 - 79)	Among whites, from ages 75 to 79 years	2.435	
	λ_{white}(80 - 84)	Among whites, from ages 80 to 84 years	2.205	
	λ_{white}(85 - 89)	Among whites, from ages 85 to 89 years	1.92	
	λ_{white}(90+)	Among whites, from ages 90+ years	1.565	

Notes: CI: Confidence Interval

Figure 1: Socsim Dementia Incidence vs Literature

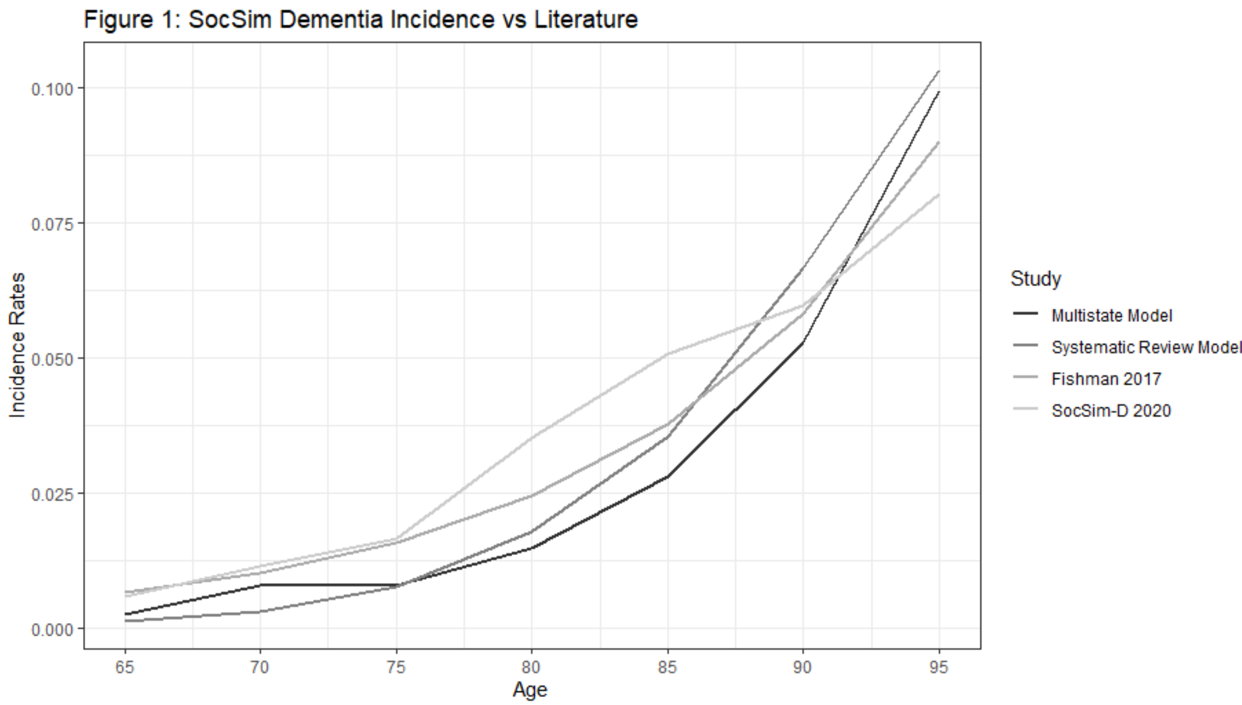


Figure 2: Dementia Survival Hazard Ratio Socsim-D Population 2000 -2060

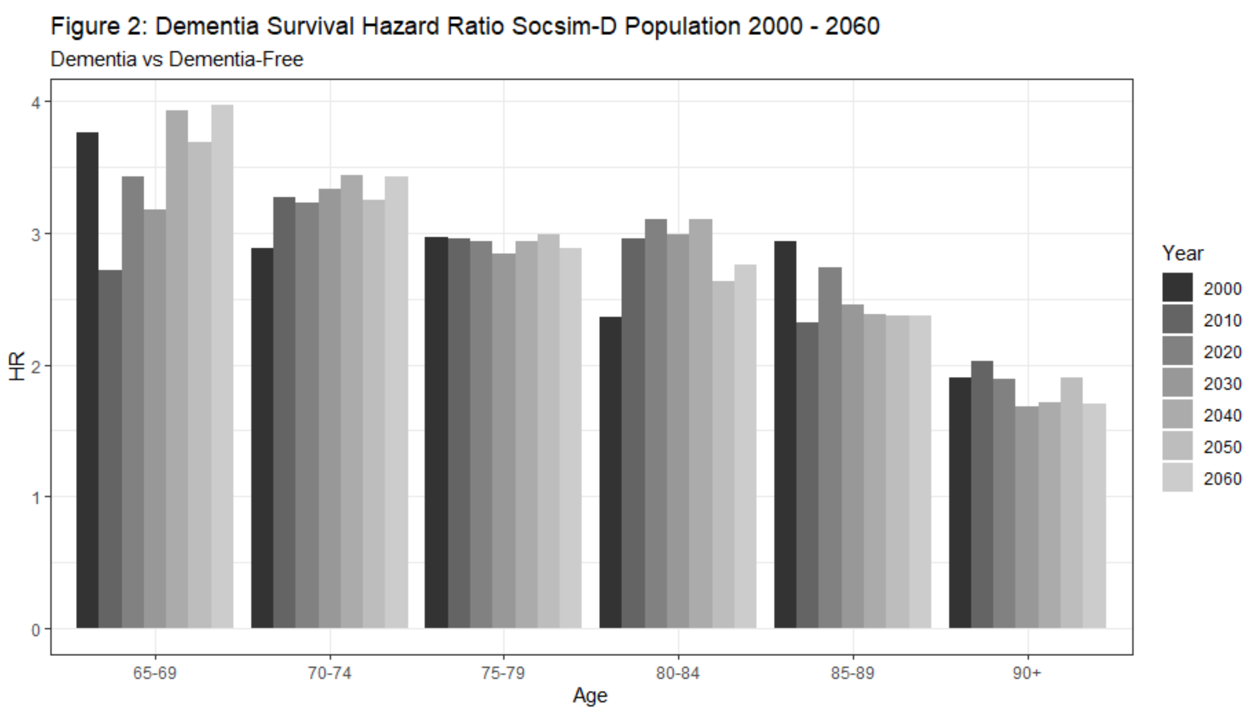
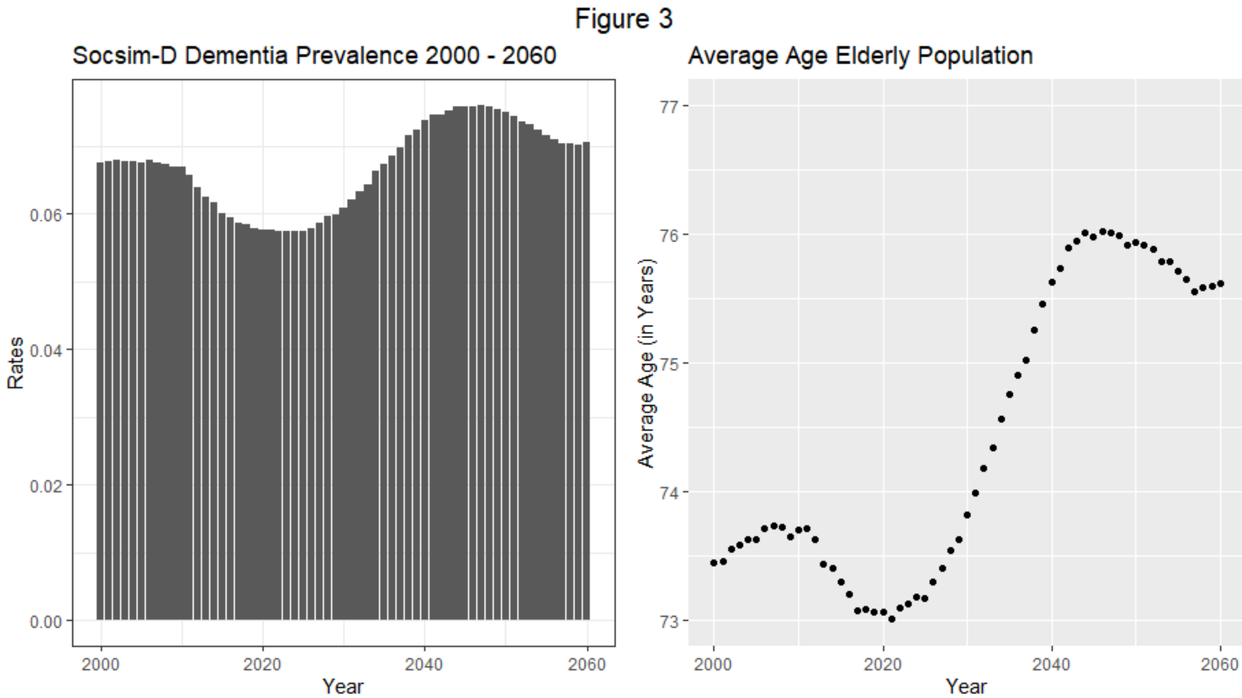


Figure 3: Average Age of the Elderly Population as an important driver of Dementia Prevalence



Appendix

Table A1: Socsim Model Calibration

	Simulated dementia prevalence by age group	Deviance (Prevalence)	Simulated years of survival after dementia diagnosis by race	Deviance [46]
Baseline				
Scenario 1	Age 65-74: 1.51% Age 75-84: 4.88% Age 85+: 10.45%	1.08	Black: 1.8 years White: 1.6 years	1.2
Scenario 2	Age 65-74: 2.73% Age 75-84: 5.56% Age 85+: 6.76%	1.05	Black: 1.9 years White: 1.7 years	1.1
Scenario 3	Age 65-74: 2.73% Age 75-84: 5.56% Age 85+: 6.76%	1.1	Black: 1.9 years White: 1.7 years	1.1
Calibration Step 1				
Scenario 1	Age 65-74: 2.74% Age 75-84: 8.51% Age 85+: 17.50%	0.31	Black: 1.83 years White: 1.75 years	0.4
Scenario 2	Age 65-74: 3.66% Age 75-84: 9.21% Age 85+: 14.43%	0.41	Black: 2.00 years White: 1.91 years	0.4
Calibration Step 2				
Scenario 1	Age 65-74: 2.48% Age 75-84: 11.02% Age 85+: 30.32%	0.06	Black: 3.25 years White: 3.33 years	0.01
Scenario 2	Age 65-74: 3.23% Age 75-84: 12.25% Age 85+: 25.73%	0.02	Black: 3.66 years White: 3.66 years	0.05

Table A2: Dementia Survival Hazard Ratio Socsim population

Year	Age	Hazard Ratio (HR)	Standard Error (HR)	Robust Standard Error (HR)	p-value
2000	65-69	3.763	0.081	0.079	0.000
2000	70-74	2.881	0.103	0.103	0.009
2000	75-79	2.972	0.099	0.097	0.015
2000	80-84	2.365	0.102	0.101	0.000
2000	85-89	2.936	0.121	0.121	0.040
2000	90+	1.905	0.167	0.171	0.000
2020	65-69	3.302	0.073	0.072	0.000
2020	70-74	3.154	0.099	0.098	0.638
2020	75-79	2.854	0.093	0.092	0.112
2020	80-84	2.799	0.091	0.090	0.067
2020	85-89	2.717	0.101	0.101	0.053
2020	90+	1.761	0.122	0.119	0.000
2030	65-69	3.158	0.078	0.079	0.000
2030	70-74	3.227	0.096	0.095	0.817
2030	75-79	2.879	0.091	0.091	0.309
2030	80-84	2.694	0.093	0.093	0.089
2030	85-89	2.542	0.099	0.101	0.031
2030	90+	1.754	0.110	0.108	0.000
2040	65-69	4.220	0.099	0.098	0.000
2040	70-74	3.163	0.116	0.115	0.013
2040	75-79	2.719	0.108	0.107	0.000
2040	80-84	3.027	0.107	0.106	0.002
2040	85-89	2.594	0.109	0.108	0.000
2040	90+	1.629	0.117	0.115	0.000
2050	65-69	3.849	0.090	0.090	0.000
2050	70-74	3.468	0.113	0.114	0.359
2050	75-79	2.908	0.103	0.103	0.006
2050	80-84	2.640	0.099	0.099	0.000
2050	85-89	2.462	0.099	0.099	0.000
2050	90+	1.935	0.101	0.100	0.000
2060	65-69	3.580	0.084	0.084	0.000
2060	70-74	3.086	0.105	0.104	0.154
2060	75-79	2.928	0.097	0.096	0.037
2060	80-84	2.780	0.095	0.094	0.007
2060	85-89	2.502	0.094	0.094	0.000
2060	90+	1.648	0.096	0.093	0.000

Table A3: Comparing Dementia Prevalence in Socsim with Published Literature

Race	Socsim (2000 - 2060)	Mayeda et al. 2017
Black	3.3	3.6
White	3.4	3.1

Chapter 3: Towards a Simple Machine Learning Tool for Predicting Dementia Status

By Ifeanyi Edochie
Pardee RAND Graduate School
January 10th, 2019

3.0 Introduction

By 2035, adults 65 and older will outnumber children less than age 18 for the first time in US history[14]. As the average age of the population rises, the prevalence of diseases such as dementia and other diseases associated with the elderly also rises. Alzheimer's disease (AD) (the most common form of dementia) is currently the 6th leading cause of death in the United States and the fifth leading cause of death among those over age 65 [15]. As the population ages and the count of dementia patients continues to grow, policymakers will need tools that assist in the modelling and projecting the cost of dementia care. The preliminary step in estimating national/regional dementia costs is in estimating the likelihood of dementia within a national representative survey of elderly using a statistical prediction model. In this study, we build on previous work[29] by using a supervised machine learning algorithm to estimate dementia status within the RAND Health and Retirement Study. Since the specificity of our model is high i.e. predicting true negatives accurately, we propose that the method could also be used as a pre-screening tool for dementia i.e. a low-cost simple tool that may be used to decide whether or not an individual should be eligible for the more sophisticated brain scan which are expensive. This could help to reduce system costs.

Regression modelling approaches have been used as the preliminary stage to projecting the cost of dementia care[29]. The previous study[29], which we build upon, uses an ordered probit regression model to classify the elderly into three groups by level of cognition: normal cognition, cognitive impairment with dementia (CIND) and dementia-positive using the RAND Health and Retirement Study (HRS). The predictors in the model include age, gender and education, measures of Activity of Daily Living[47] and Instrumental Activity of Daily Living (IADL) limitations and lagged cognitive functioning measures. With their model, 89.8% of non-demented patients were classified correctly (specificity) while 77.9% of the dementia-positive were classified correctly. Overall, the model predicts 85% of cases correctly[29]. However, since their analysis is simply concerned with estimating overall costs, the predictive power of their approach is never tested against an untrained sample. Therefore, the true accuracy of their model remains unknown. As is standard in the machine learning field, our model is trained on 75% of the data so that the remainder is utilized to test the predictive power of our analysis.

The application of machine learning (ML) methods in the medical domain has become more prominent in recent years. ML techniques such as the multiple instance learning method were used to detect morphological abnormalities from brain MRI scans to assist dementia diagnosis[48]. Starting with MRIs of 834 patients the study achieves a classification accuracy of 89% amongst Alzheimer's and non-Alzheimer's patients. Other studies have used image-based biomarkers for AD with similar classification accuracy and specificity rates[49-51]. However, brain MRI scans for patients can cost thousands of dollars which can be expensive. Our approach proposes a low-cost complement which could serve as a preliminary screening tool before a more formal diagnosis is made. This could be important for reducing costs for the medical insurer.

Machine learning techniques have been used on cognitive function data to predict whether a patient has any type of cognitive impairment (which includes dementia). A study[52] uses ML methods to estimate the best decision rules to distinguish a brain with normal cognition from a brain with dementia using the functional activities questionnaire as well as the Mini-Mental Status Exam i.e. a combination cognitive and non-cognitive measures. The classification accuracy of their model is 93% for sensitivity and 80% for specificity. Although, our classification model predicts dementia status less accurately it performs comparably (specificity = 87%, sensitivity = 72%).

Another study[53] uses unsupervised machine learning to identify high likelihood of dementia within the RAND Health and Retirement Study (HRS). Their research combines demographic characteristics with the activities of daily living variables as well as other cognitive measures. Their work revealed three clusters. One for respondents "without any functional or motor limitations", cluster 2 with walking/climbing limitations and the final cluster with both functional and walking/climbing limitations. When cluster 3 was compared with the predicted probabilities of dementia in the HRS, they corresponded to the high probability dementia cases. However, the unsupervised learning algorithm operates as a black box. Since the sole purpose of the analysis is to predict outcomes accurately, some may consider this unproblematic. However, developing an understanding of what factors drive the predicted outcomes is also important for building more accurate models in the future. Unsupervised learning algorithms are unable to provide this information intuitively. This is the advantage that certain supervised learning algorithms will have over the unsupervised models.

In this study, I use simple supervised learning algorithm to predict dementia cases within the HRS using only demographic factors and the activities of daily living[47] as well as indicators for known pre-existing conditions. Our analysis does not include any measures of cognition. This is an important distinction because measures of cognition might not always be readily available. Part of the motivation behind this modelling exercise seeks to test how accurately dementia status can be estimated without cognitive measures. I show that without measures of cognition a supervised learning algorithm still performs comparably to previous work mentioned here.

The primary purpose of our analysis is to test how accurately (in terms of specificity and sensitivity) a simple supervised ML model can detect dementia status using the demographic factors, ADLs and indicators for pre-existing medical conditions. This machine learning approach for predicting dementia status could form the basis for future predictions on the cost of dementia in the US. This paper has certain advantages over the previously aforementioned studies. I attempt to provide a lower cost preliminary screening mechanism for dementia status so that individuals flagged as positive may then go through the formal testing process, which is typically costlier. Therefore, a multilayer perceptron algorithm (a type of neural network model) is used to classify individuals by dementia status.

3.1 Data Description

3.1.1 Survey Samples

The HRS is a longitudinal population-based household survey conducted by the Institute for Social Research at the University of Michigan sponsored by the National Institute on Aging (NIA U01AG009740) and the Social Security Administration. The survey collects information on the economic well-being and health status of Americans over the age of 50 and is nationally representative. The first cohort was surveyed in 1992 and subsequent iterations have occurred biennially, replacing individuals who die with younger entrants. There are currently 12 waves of the survey. The HRS sample is stratified geographically and covers all demographic groups. Each wave of the HRS sample contains approximately 20,000 respondents who are randomly selected from the set of all eligible household members. The data is collected in the form of interviews. There are face-to-face initial interviews at first contact of respondents with the survey team. We chose the full set of panel data from wave 10 through 12 (from 2000 to 2012, ~55000 observations) for the training of our model since these are the only waves in which dementia information is collected within the survey.

3.1.2 Measures

For the analysis we used variables for demographics, pre-existing medical conditions, labor market participation variables and activities of daily living from the HRS. The demographic factors used in our model include indicators of sex, race, educational attainment and age. For each activity of daily living (ADL), there is a binary variable for whether an individual can perform that activity with some difficulty (0 = no difficulty; 1 = difficulty). In addition, each respondent is also asked to rate how well they can perform that activity on a scale of 1 to 5. The activities of daily living seek to understand whether individuals can perform the following activities include bathing, dressing, eating, getting into and out of bed as well as walking. In addition, it also seeks to understand how well (subjectively speaking) these activities are performed on a scale of 1 to 5. These variables are combined to form a composite index. Another set of the variables, the instrumental activities of daily living (IADL) measure ability to carry out other functional tasks which could be useful measures and they were constructed within the HRS to provide consistency across waves. There are 5 IADLs for measuring ability to: use the telephone, ingest medication, grocery shopping, cooking and handling money. We also leverage a key advantage of the HRS in using its subjective measures of labor choice. The HRS asks respondents on a scale of 1 to 100, how likely they think they are to work past the ages of 62 and 65 respectively. In addition, we collect information on the functional activities performed in employment such as lifting, stooping, seeing etc. The final set of independent factors provide information on the pre-existing medical conditions of respondents including whether they have/had strokes, heart diseases, lung problems, cancers, diabetes and high blood pressure issues.

All these variables are easy to collect from individuals and are therefore of great advantage if they can be used in classifying respondents by dementia status.

We seek to build a model that focuses on variables that can easily be collected from prospective respondents. The outcome is an indicator variable measuring whether a respondent has been doctor diagnosed with dementia. In all, there are 54 independent variables used in our analysis with an outcome variable for dementia status. We have included summary statistics for the HRS data used in our analysis in the appendix section in Table A1.

3.2 Methods

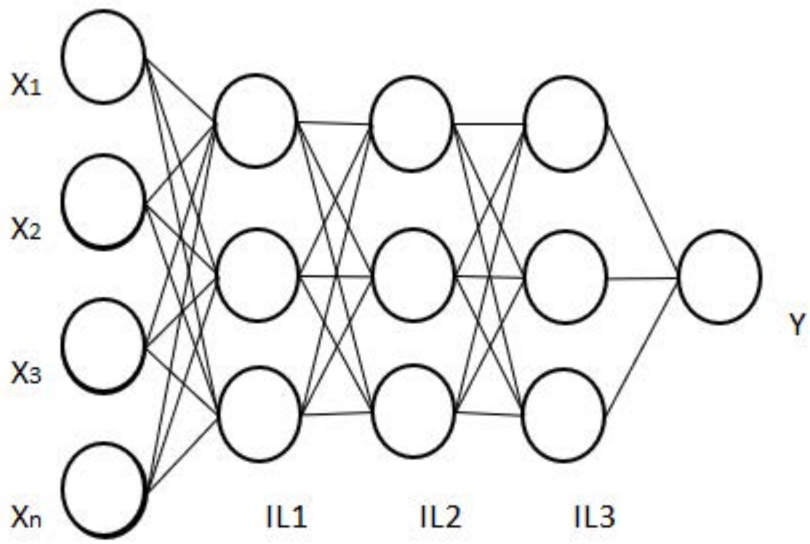
3.2.1 Supervised Machine Learning Classification

We run a neural network, specifically, a multilayer perceptron (MLP) model, to predict dementia status in the HRS using the 54 variables previously described as the predictors. The analysis uses R's Stuttgart Neural Network Simulator (RSNNS). The major advantage of using this package over other R packages are its speed of convergence as well as its ability to implement a multi-layer perceptron quickly.

A perceptron is a linear or non-linear classifier (algorithm) that produces a single output, y, based on a vector, X, real-valued inputs by forming a linear combination of variables and weights.

$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \varphi(W^T X + b)$$

In the above equation, w denotes weights, b is the bias term and phi is the activation function. An MLP is a deep, artificial neural network composed of more than a single perceptron. A perceptron is the basic unit of a neural network containing just a single node. The perceptron consists of inputs (i.e. like input variables of a regression, x_i) and their estimated weights (i.e. like the coefficients in a regression, w_i). The single perceptron with a linear activation is the least square regression model. A linear activation function simply multiplies these weights by the inputs to estimate the output, y, which is the used in the classification i.e. dementia status in our case. A multilayer perceptron is made up of at least 3 nodes which are arranged in layers. The input layer (X_1 - X_n in figure 1 below) contained all variables as previously described. An MLP has at least one hidden layer. The model used in this analysis contains three because any more layers came at a large computational time cost without a significant increase in classification accuracy. Each inner layer (IL1, IL2 and IL3 as in the Figure 1) is an interaction of the previous layer of variables. The arrows in figure 1 show the nature of interactions used in the model. The activation function is a sigmoid logistic function which transforms a linear combination of weights and all variables into a number between 0 and 1 i.e. a probability output, Y.



Diagramatic Description: Multilayer Perceptron visualization

The task of this model framework is to train on the HRS data described by adjusting the parameters (weights and biases) towards minimizing the root mean squared error.

3.2.2 MLP Process

In estimating the machine learning algorithm, we separate the data into a training set and a test set. The training set is a 75% random sample of the HRS data while the remainder of the sample is used as the test set to ascertain the quality of model prediction. We rescale all independent variables to real-values between 0 and 1 as machine learning algorithms fail to converge on a solution set with large values. Using the HRS training set, we train the model according to the following relation:

$\hat{y} = \varphi(W^T X + b)$

We project the training model onto the test data to predict dementia status. We compute estimates for the specificity and sensitivity of the model. However, given that dementia positivity is much less common within the dataset, we have a sample imbalance problem. Within the HRS, there are 38 non-dementia cases to each dementia case which makes predicting a case of dementia better than a 50-50 chance highly likely. To address this problem, we use the Synthetic Minority Over-Sampling Technique (SMOTE) [54]. This technique synthetically increases the size of the minority class observations to maximize prediction accuracy. We also present confusion matrices and other metrics to measure the contribution of each variable to the quality of model predictions. This is an advantage of our modelling approach over the unsupervised learning approaches that have been used to address this classification problem.

3.3 Results

3.3.1 Baseline Model Results

First, we run a simple MLP model with 3 hidden layers with standard backpropagation as the learning function to train the 75% sample of the RAND HRS data (n = 38132) previously described. The analysis results in a specificity of 0.98 and a sensitivity of 0.78 for the test set. Below is the confusion matrix (i.e. a false positive/false negative table) on the training set:

Table 3.1: Confusion Matrix for Training Data (Baseline Model)

	Non-Dementia	Dementia
Predicted Non-Dementia	0.98	0.02
Predicted Dementia	0.22	0.78

We use the same model to predict dementia status on the HRS test set (n = 13020). The results are as follows:

Table 3.2: Confusion Matrix for Test Data (Baseline Model)

	Non-Dementia	Dementia
Predicted Non-Dementia	0.98	0.02
Predicted Dementia	0.64	0.36

It is evident from the analysis that since dementia is a rear event within the HRS population (3% dementia cases), the model predicts non-dementia cases more easily than the smaller class of dementia cases. Therefore, each dementia case has a 64% chance of being predicted as not having dementia.

3.3.2 Fixing the Class Imbalance Problem

There are 38 non-dementia cases for each case of dementia in the relevant waves of the dataset. We apply the Synthetic Minority Over-Sampling Technique[54] (SMOTE) by synthetically replicating each dementia observation 37 times so that our training sample now is a balanced dataset for training a model to predict dementia status. The training sample now contains 73578 observations. The confusion matrix resulting from a re-training is as follows:

Table 3.3: Confusion Matrix for Training Data (SMOTE'd Model)

	Non-Dementia	Dementia
Predicted Non-Dementia	0.87	0.13
Predicted Dementia	0.04	0.96

The balanced data reduces the specificity of the model (0.98 to 0.78) but there are obvious gains in the sensitivity of the model (0.78 to 0.96). The new MLP neural network model is used to predict dementia status in the same set of test cases. The results are as follows:

Table 3.4: Confusion Matrix for Test Data (SMOTE'd Model)

	Non-Dementia	Dementia
Predicted Non-Dementia	0.87	0.13
Predicted Dementia	0.28	0.72

By attempting to fix the class imbalance problem, the neural network predicts dementia positivity with a sensitivity of 0.72 (up from 0.36). The sensitivity of the minority class is significantly larger than 50% which means model sensitivity is a 22% improvement on a simple coin toss guess.

3.3.3 Model Interpretation

One of the major criticisms of machine learning operations is in the interpretability of the results. Often times models operate in a black box environment in which high accuracy has been achieved without an understanding of how specific model inputs lead to outputs. In this analysis, we use Garson's algorithm[55] to determine which variables contribute the most to the results. Please find the charts in the Appendix Figure A1.

The weights applied to the variables in the deep learning MLP model are in part analogous to parameter coefficients in a standard regression. The difference in interpretation however is that the Gauss Markov assumptions behind regression models will not apply to our deep learning MLP model. We use Garson's algorithm[55] to show the relative importance of the predictor variables for predicting dementia status within our MLP model. This is determined by identifying all weighted connections between the nodes of interest i.e. all the weights connected to the specific input node that pass through the hidden layer to the dementia status variable are identified and repeated for each of the 54 predictors. Each weight associated to a predictor is tallied and scaled relative to all other inputs.

From the results in the appendix Figure A1, we see that the composite variable for instrumental activities of daily living is the most significant variable in predicting the disease status. This is followed by the age variable which is as expected since age is understood to be the strong predictor of dementia status. We also find that pre-existing spousal medical conditions, are the next important variable as well.

3.4 Discussion

Our approach is not without its limitations. Dementia information within the HRS is collected through doctor diagnosis. Since doctor diagnosis is a function of health care access, individuals within lower socioeconomic status groups are less likely to be screened for dementia. Hence, the model as currently trained will likely incorrectly predict dementia status amongst individuals from lower SES groups.

The analysis in this study adds to the literature on developing predictive models for estimating dementia status. Specifically, I create a predictive model to estimate cost of dementia care. The model can also be deployed as a pre-screening tool that predicts whether HRS respondents have dementia without using any measures of cognition. We use a deep learning neural network model (multilayer perceptron) to predict dementia status amongst HRS respondents. Since only 3 percent of the observations are dementia positive, we have a severe class imbalance problem which makes a baseline prediction model more likely to lack statistical specificity i.e. predicting the dementia cases as having dementia. To deal with this problem, we use the Synthetic Minority Over-Sampling Technique (SMOTE)[54] to increase minority class representation. This technique doubles the specificity (36% to 72%) while maintain the sensitivity of the prediction process. We are also able to show that the composite instrumental activities of daily living variable, as well as age and the pre-existing medical conditions for the spouse are the best predictors of dementia status.

Appendix

Figure B1: Garson's Algorithm for Relative Importance of MLP model inputs

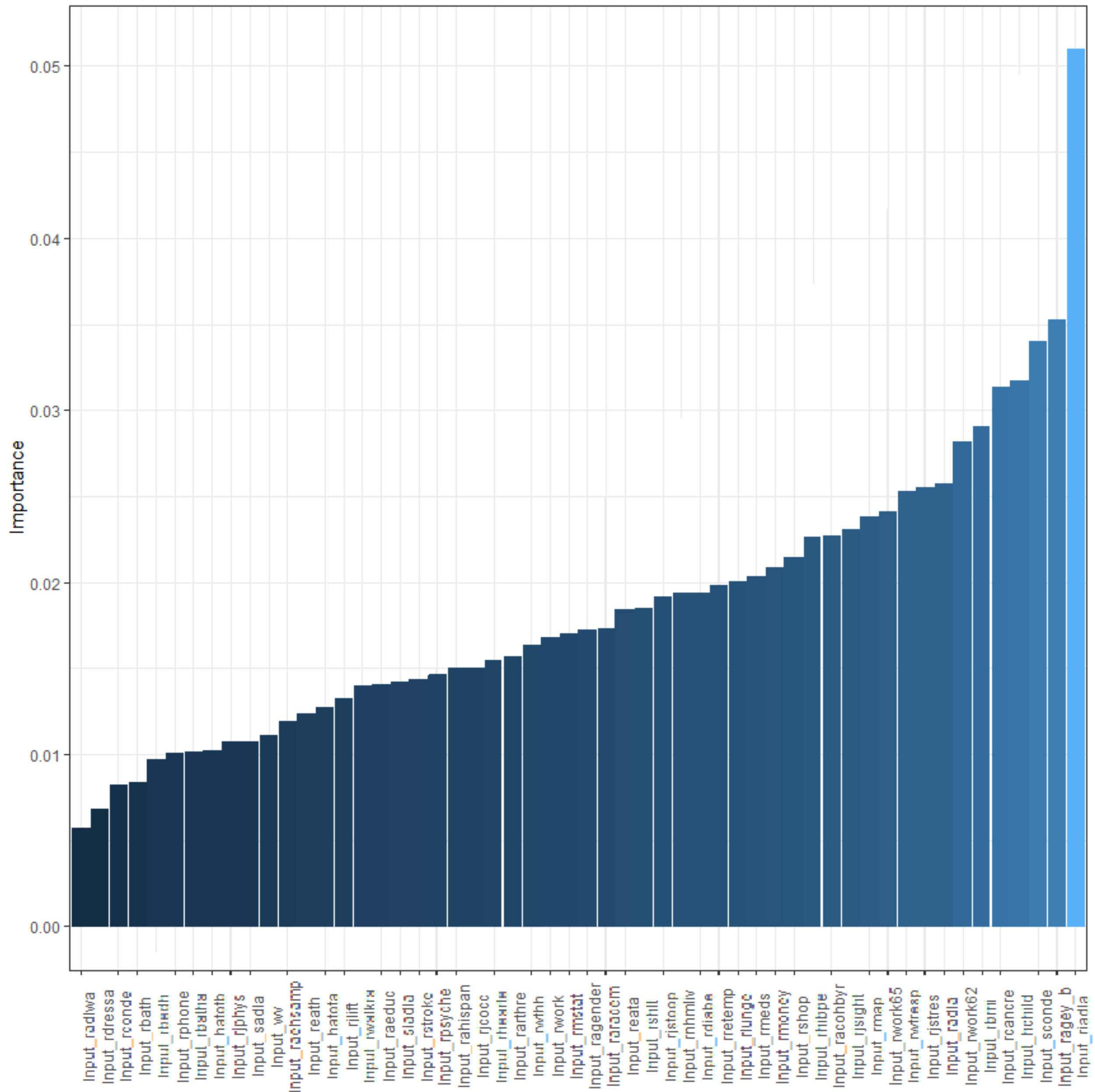


Table B1: Descriptive Tables

Variable	Dementia Negative	Dementia Positive	P-Value
wv	10.99	11.05	0.42
rmstat	3.11	4.27	0.00
ragender	1.57	1.60	0.29
rahispan	1.13	1.13	0.50
raracem	1.33	1.28	0.07
raeduc	3.26	2.81	0.00
ragey_b	69.01	78.70	0.00
raehsamp	1.49	1.48	0.38
rwthh	4324.05	2788.67	0.00
rwtresp	4728.18	3001.21	0.00
racohbyr	4.93	3.75	0.00
sconde	2.07	2.32	0.00
sadla	0.17	0.25	0.00
siadla	0.08	0.11	0.00
rshlt	2.89	3.81	0.00
rbmi	28.43	26.38	0.00
rhibpe	1.63	1.77	0.00
rdiabe	1.25	1.33	0.02
rcancre	1.16	1.22	0.05
rlunge	1.10	1.20	0.00
rhearte	1.26	1.48	0.00
rstroke	1.09	1.39	0.00
rpsyche	1.18	1.47	0.00
rarthre	1.61	1.75	0.00
ralzhe	1.00	1.02	0.29
rdemen	1.01	2.00	0.00
rconde	2.28	3.60	0.00
rbath	0.08	0.59	0.00
reath	0.01	0.24	0.00
rbedh	0.03	0.30	0.00
rshop	1.16	2.03	0.00
rphone	0.06	0.76	0.00
rmeds	0.04	0.46	0.00
rmoney	0.32	1.32	0.00
rmap	0.76	2.72	0.00
rdressa	1.11	1.47	0.00
rwalkra	1.07	1.42	0.00
rbatha	1.07	1.51	0.00
reata	1.03	1.32	0.00

radlwa	0.21	1.31	0.00
radla	0.35	2.09	0.00
riadla	0.13	1.53	0.00
rnhmliv	1.01	1.26	0.00
hatota	420132.11	309761.87	0.00
rwork62	40.65	37.06	0.00
rwork65	16.86	9.58	0.00
rretemp	1.59	1.78	0.00
rwork	1.35	1.03	0.00
rjphys	2.94	2.99	0.26
rjlift	3.80	3.98	0.05
rjstoop	2.99	2.99	0.50
rjsight	1.17	1.02	0.00
rjstres	2.13	2.01	0.01
rjcocc	3.26	3.06	0.01
hchild	3.17	3.26	0.16

Conclusion

Age dependency is a rising phenomenon amongst many countries. They generally lie in two distinct categories. The first set of countries have higher fertility rates and tend to have a larger population of young i.e. less than 18 years of age. Many of these countries are in Africa (with sometimes half the population under age 15) and to a lesser extent also in Latin America and the Caribbean Islands. The other group have lower fertility rates coupled with advancements in medical technology that keep the elderly living longer. As a result, these populations are aging i.e. high prevalence of individuals over age 65. These distinct population phenomena are both commensurate with shrinking populations that are generally responsible for catering for those who are too young to work enough hours to cater for themselves as well as the elderly who ideally should be retired. Analytical techniques could play important roles in assisting policymakers in creating futuristic plans for these populations.

In this dissertation, we apply a variety of predictive modelling approaches to address three specific issues within the scope of the phenomena previously described. In the first paper, I use a microsimulation approach to estimate the effects of conditional and unconditional cash transfer programs on school enrollment and poverty reduction for children between the ages of 7 and 15. This is based on the Nigerian government opting for an unconditional cash transfer program in 2017 haven being unable to run a conditional cash transfer program as planned in 2017. Our analysis shows that for the same transfer amount the unconditional cash transfer program yields far lower increments in educational enrollment than a conditional program. However, we should that depending on the program design that the government could choose to invest in monitoring a conditional program between 1.8% and 27.5% of the costs of a UCT depending on the program design.

The second study uses develops a novel microsimulation approach to project the prevalence of dementia up till 2060. We assign dementia status to a synthetic population of agents that matches the elderly population of the United States according to the epidemiology of dementia. We pull dementia information from a wide range of robust studies to inform parameters which we calibrate within our model. One of the important features of our modelling approach is our ability to simulate dementia incidence without making any changes the distribution of mortality within the synthetic population. We find that age (not race or gender) will be the predominant driver of dementia prevalence in the United States. Our results show a current declining trend in the prevalence rate of dementia which begins to increase after 2024 up till the mid-2050s where the rate becomes relatively level.

In the final paper, I use a machine learning approach to predict dementia status within a dataset like the RAND Health and Retirement Study without using any measures of cognition. This is important for two reasons. On one hand, projecting the costs of dementia care typically requires regression-based models from which future cost projections are made. Secondly, our model specificity and sensitivity is 0.87 and 0.72 in out sample predictions, i.e. dementia negativity correctly 87% of the time while predicting dementia positive cases correctly 72% of the time. This could be an important pre-screening tool for the elderly to reduce system costs. Since brain scans can be expensive, an approach that can screen out individuals without dementia could reduce system costs since nearly 9 out of 10 times a patient without dementia is likely to be correctly screened out.

Discussion

Governments in countries with high young age dependency ratios particularly in African countries have long used cash transfers for poverty reduction as well as to increase child school enrollment. This brings up an important discussion about the nature of work for children. African governments as well as the international development community have long debated whether children should work, the nature of child labor if permitted and the dynamics of the work and school for children. As these cash transfer programs that seek to increase educational enrollment become more prevalent, African governments need to define clearer child labor policy. It is important that these countries continue to develop country-wide models with enough sophistication to evaluate scenarios for their programs.

Western countries have a different challenge. Policymakers will have to grapple with the rising prevalence of age-old diseases amongst the overall population. Many developed countries with the rising tide of low fertility and other demographic concerns have begun adopting more open immigration policies. Canada as well as some many European countries have used created paths to citizenships to usher in younger demographics into their countries. While this is multifaceted political issue many countries, policymakers have this option is a means of providing caregiving to the large cohort of elderly who might go with access to the family care.

Bibliography

1. Paul Schultz, T., *School subsidies for the poor: evaluating the Mexican Progresa poverty program.* Journal of Development Economics, 2004. **74**(1): p. 199-250.
2. Fiszbein, A. and N.R. Schady, *Conditional cash transfers: reducing present and future poverty*. 2009: World Bank Publications.
3. Kharas, H.H., Kristofer; Hofer, Martin, *The start of a new poverty narrative*, in *Future Development*. 2018, Brookings: Washington D.C.
4. Jones, H., *More Education, Better Jobs? A Critical Review of CCTs and Brazil's Bolsa Familia Programme for Long-Term Poverty Reduction.* Social Policy and Society, 2016. **15**(3): p. 465-478.
5. Parker, S.W., L. Rubalcava, and G. Teruel, *Chapter 62 Evaluating Conditional Schooling and Health Programs*, in *Handbook of Development Economics*, T.P. Schultz and J.A. Strauss, Editors. 2007, Elsevier. p. 3963-4035.
6. Sudhanshu, H. and D. Benjamin, *The Experience of Conditional Cash Transfers in Latin America and the Caribbean.* Development Policy Review, 2006. **24**(5): p. 513-536.
7. Alviar, C., et al., *The impact of Kenya's Cash Transfer for Orphans and Vulnerable Children on human capital.* Journal of Development Effectiveness, 2012. **4**(1): p. 38-49.
8. Bourguignon, F., F.H.G. Ferreira, and P.G. Leite, *Conditional cash transfers, schooling, and child labor: Micro-simulating Brazil's Bolsa Escola program.* World Bank Economic Review, 2003. **17**(2): p. 229-254.
9. Mideros, A., F. Gassmann, and P. Mohnen, *Estimation of rates of return on social protection: ex ante microsimulation of social transfers in Cambodia.* Journal of Development Effectiveness, 2016. **8**(1): p. 67-86.
10. Peruffo, M. and P.C. Ferreira, *THE LONG-TERM EFFECTS OF CONDITIONAL CASH TRANSFERS ON CHILD LABOR AND SCHOOL ENROLLMENT.* Economic Inquiry, 2017. **55**(4): p. 2008-2030.
11. Samson, M., *Micro-simulation analysis of social protection interventions in Pacific Island countries.* AusAID, Canberra, Australia, 2012.
12. Catubig, M.C.L. and R.A. Villano, *Conditional Cash Transfer and School Outcomes: An Evaluation of the Pantawid Pamilyang Pilipino Program in Davao Oriental, Philippines.* Asian Economic Journal, 2017. **31**(4): p. 403-421.
13. Bhalla, G., et al., *The effect of cash transfers and household vulnerability on food security in Zimbabwe.* Food Policy, 2018. **74**: p. 82-99.
14. US Census Bureau, *Fertility of Women in the United States: 2018*. 2018.
15. Mayeda, E.R., et al., *Survival after dementia diagnosis in five racial/ethnic groups.* Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 2017. **13**(7): p. 761-769.
16. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders, 4th ed., text rev. (DSMIV-TR).* . 2000, Arlington, VA.
17. Langa, K.M., et al., *National estimates of the quantity and cost of informal caregiving for the elderly with dementia*.* Journal of general internal medicine, 2001. **16**(11): p. 770-778.

18. Bookman, A. and D. Kimbrel, *Families and elder care in the twenty-first century.* The Future of Children, 2011. **21**(2): p. 117-140.
19. Alzheimer's Association, *2016 Alzheimer's disease facts and figures.* Alzheimer's & Dementia, 2016. **12**(4): p. 459-509.
20. Hamilton, B.E., et al., *Births: provisional data for 2018.* 2019.
21. Rutter, C.M. and J.E. Savarino, *An Evidence-Based Microsimulation Model for Colorectal Cancer: Validation and Application.* Cancer Epidemiology Biomarkers & Prevention, 2010. **19**(8): p. 1992-2002.
22. Mazzaferro, C. and M. Morciano, *CAPP_DYN: A Dinamic Microsimulation Model for the Italian Social Security System*. 2008: Università degli Studi di Modena e Reggio Emilia, Dipartimento di Economia ….
23. Laditka, S.B. and D.A. Wolf, *New Methods for Analyzing Active Life Expectancy.* Journal of Aging and Health, 1998. **10**(2): p. 214-241.
24. Martini, A. and U. Trivellato, *The role of survey data in microsimulation models for social policy analysis.* Labour, 1997. **11**(1): p. 83-112.
25. SOCSIM. 2016 [cited 2016 April 15]; Available from: http://lab.demog.berkeley.edu/socsim/.
26. Zagheni, E., *Microsimulation in demographic research.* International encyclopedia of social and behavioral sciences, 2015. **2**: p. 343-346.
27. Brookmeyer, R., et al., *Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States.* Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 2018. **14**(2): p. 121-129.
28. RAND, *RAND HRS Data, Version N*. 2014, RAND.
29. Hurd, M.D., et al., *Supplementary Appendix: Monetary Costs of Dementia in the United States.* 2013.
30. Heron, M., *Deaths: Leading causes for 2014.* National Vital Statistics Report, 2016. **65**(5).
31. Knopman, D.S., et al., *Survival study of vascular dementia in rochester, minnesota.* Archives of Neurology, 2003. **60**(1): p. 85-90.
32. Ganguli, M., et al., *Alzheimer disease and mortality: A 15-year epidemiological study.* Archives of Neurology, 2005. **62**(5): p. 779-784.
33. Helzner, E.P., et al., *Survival in Alzheimer disease: A multiethnic, population-based study of incident cases.* Neurology, 2008. **71**(19): p. 1489-1495.
34. Wilson, R.S., et al., *Biracial population study of mortality in mild cognitive impairment and alzheimer disease.* Archives of Neurology, 2009. **66**(6): p. 767-772.
35. Little, D.M., et al., *Mortality, Dementia, and Apolipoprotein E Genotype in Elderly White Women in the United States.* Journal of the American Geriatrics Society, 2009. **57**(2): p. 231-236.
36. James, B.D., et al., *Contribution of Alzheimer disease to mortality in the United States.* Neurology, 2014. **82**(12): p. 1045-1050.
37. Breslow, N.E. and N.E. Day, *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies*. Vol. 1. 1980: Distributed for IARC by WHO, Geneva, Switzerland.
38. Evans, D.A., *Estimated Prevalence of Alzheimer's Disease in the United States.* The Milbank Quarterly, 1990. **68**(2): p. 267-289.

39. Hebert, L.E., et al., *Alzheimer disease in the United States (2010–2050) estimated using the 2010 census.* Neurology, 2013. **80**(19): p. 1778-1783.
40. Hudomiet, P., M.D. Hurd, and S. Rohwedder, *Dementia Prevalence in the United States in 2000 and 2012: Estimates Based on a Nationally Representative Study.* The Journals of Gerontology: Series B, 2018. **73**(suppl_1): p. S10-S19.
41. Koller, D. and J.P.W. Bynum, *Dementia in the USA: state variation in prevalence.* Journal of Public Health, 2015. **37**(4): p. 597-604.
42. Langa, K.M., et al., *A Comparison of the Prevalence of Dementia in the United States in 2000 and 2012.* JAMA Internal Medicine, 2017. **177**(1): p. 51-58.
43. Brookmeyer, R. and N. Abdalla, *Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease.* Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 2018. **14**(8): p. 981-988.
44. Fishman, E., *Risk of Developing Dementia at Older Ages in the United States.* Demography, 2017. **54**(5): p. 1897-1919.
45. Cobb, J.L., et al., *The effect of education on the incidence of dementia and Alzheimer's disease in the Framingham Study.* Neurology, 1995. **45**(9): p. 1707-1712.
46. Hanley, J.A., *Nelson–Aalen Estimator.*
47. Larson, E.B., et al., *SUrvival after initial diagnosis of alzheimer disease.* Annals of Internal Medicine, 2004. **140**(7): p. 501-509.
48. Tong, T., et al., *Multiple instance learning for classification of dementia in brain MRI.* Medical Image Analysis, 2014. **18**(5): p. 808-818.
49. Wolz, R., et al., *Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease.* PLOS ONE, 2011. **6**(10): p. e25446.
50. Coupé, P., et al., *Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease.* NeuroImage: Clinical, 2012. **1**(1): p. 141-152.
51. Liu, M., et al., *Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis.* 2014. **35**(4): p. 1305-1319.
52. Shankle, W.R., et al. *Detecting very early stages of dementia from normal aging with Machine Learning methods*. 1997. Berlin, Heidelberg: Springer Berlin Heidelberg.
53. Cleret de Langavant, L., E. Bayen, and K. Yaffe, *Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study.* Journal of medical Internet research, 2018. **20**(7): p. e10493-e10493.
54. Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique.* Journal of artificial intelligence research, 2002. **16**: p. 321-357.
55. Garson, G.D., *Interpreting neural-network connection weights.* 1991. **6**(4): p. 46-51.