Modern correlates of health

ECON 40565
Health Economics
Fall 2020

Introduction

• Most of this class we will examine markets for medical care
  – How they operate
  – What are economic issues
• Medical care is however only interesting in that it is an intermediate product – used to produce what people care about – health
• This section – discuss what inputs can be transformed into health outputs

Three main issues

• How is health measured?

• Some predictors of outcomes?

• Extended discussion about the role of socioeconomic status and health

Aggregate measures of health

• Mortality rates
  – death per period among a define population

• Infant mortality rate
  – deaths 1st year of life/births
  – Neonatal mortality: deaths 1st 28 days

• Life expectancy
  – At birth
  – Conditional on a particular age
Self-reported health status

- **Benefits**
  - Easy/low cost variable to collect
  - Predicts other measures of health that are difficult to collect

- **Shortcomings**
  - No way to compare people
  - No way to compare aggregate data across countries
  - May be difficult to compare groups over time
    - Rise in disability
    - “Harvesting”

---

% Reporting Health Status, Males

<table>
<thead>
<tr>
<th>Health</th>
<th>Age 30-44</th>
<th>Age 45-64</th>
<th>Age 65-74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>43.7%</td>
<td>30.6%</td>
<td>18.1%</td>
</tr>
<tr>
<td>Very good</td>
<td>30.3%</td>
<td>26.9%</td>
<td>22.5%</td>
</tr>
<tr>
<td>Good</td>
<td>19.8%</td>
<td>26.1%</td>
<td>31.6%</td>
</tr>
<tr>
<td>Fair</td>
<td>4.7%</td>
<td>10.6%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Poor</td>
<td>1.5%</td>
<td>5.8%</td>
<td>9.3%</td>
</tr>
</tbody>
</table>

---

5-Year Mortality Rate, Males

<table>
<thead>
<tr>
<th>Health</th>
<th>Age 30-44</th>
<th>Age 45-64</th>
<th>Age 65-74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>0.7%</td>
<td>2.4%</td>
<td>8.6%</td>
</tr>
<tr>
<td>Very good</td>
<td>0.9%</td>
<td>2.9%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Good</td>
<td>1.6%</td>
<td>5.2%</td>
<td>16.7%</td>
</tr>
<tr>
<td>Fair</td>
<td>2.9%</td>
<td>11.7%</td>
<td>25.2%</td>
</tr>
<tr>
<td>Poor</td>
<td>10.4%</td>
<td>22.8%</td>
<td>42.9%</td>
</tr>
</tbody>
</table>

---

% Reporting Health Status, Females

<table>
<thead>
<tr>
<th>Health</th>
<th>Age 30-44</th>
<th>Age 45-64</th>
<th>Age 65-74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>0.3%</td>
<td>1.7%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Very good</td>
<td>0.4%</td>
<td>1.9%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Good</td>
<td>0.9%</td>
<td>2.9%</td>
<td>8.8%</td>
</tr>
<tr>
<td>Fair</td>
<td>1.8%</td>
<td>6.2%</td>
<td>14.1%</td>
</tr>
<tr>
<td>Poor</td>
<td>7.1%</td>
<td>15.6%</td>
<td>32.2%</td>
</tr>
</tbody>
</table>

---

5-Year Mortality Rate, Females

<table>
<thead>
<tr>
<th>Health</th>
<th>Age 30-44</th>
<th>Age 45-64</th>
<th>Age 65-74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>0.3%</td>
<td>1.7%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Very good</td>
<td>0.4%</td>
<td>1.9%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Good</td>
<td>0.9%</td>
<td>2.9%</td>
<td>8.8%</td>
</tr>
<tr>
<td>Fair</td>
<td>1.8%</td>
<td>6.2%</td>
<td>14.1%</td>
</tr>
<tr>
<td>Poor</td>
<td>7.1%</td>
<td>15.6%</td>
<td>32.2%</td>
</tr>
</tbody>
</table>
Biomarkers

- Mortality limited for some populations
- SRHS difficult to compare across people
- Objective way to measure health status across people?
- Biomarkers
  - Clinical markers of physiology
  - Predictive of future health outcomes
  - Measurable across people
  - Easily collect

Examples

- Blood pressure
  - High BP can lead to stroke, AMI, heart failure, kidney failure
- Cholesterol
  - HDL, LDL and total
  - High chol. can lead to heart attack
- Resting heart rate
- Glycated hemoglobin
  - Predictor of diabetes,
- Body mass index (kg's/cm²)
  - Increased risk of diabetes
  - High BMI correlated w/ increased mortality

Mortality rates in the 20th century

- Tremendous changes in aggregate statistics
- Two halves
  - Decline in infant deaths (1/2 half) and infections
  - Conquering cardiac disease
What causes big changes in life expectancy?

- Most deaths are to the elderly
- But, when an infant dies, you add a small number to the numerator in a life expectancy calculation
- Big changes will be generated by
  - Changes in the infant mortality rate
  - Changes in mortality for the elderly which are a large fraction of deaths

**Distribution of Deaths by Age**

<table>
<thead>
<tr>
<th>Age of deaths</th>
<th>Fraction of deaths</th>
<th>Age of deaths</th>
<th>Fraction of deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td>1.0%</td>
<td>55-64</td>
<td>12.9%</td>
</tr>
<tr>
<td>1-14</td>
<td>0.3%</td>
<td>65-74</td>
<td>16.5%</td>
</tr>
<tr>
<td>15-24</td>
<td>1.1%</td>
<td>75-84</td>
<td>24.9%</td>
</tr>
<tr>
<td>25-34</td>
<td>1.7%</td>
<td>85+</td>
<td>31.3%</td>
</tr>
<tr>
<td>35-44</td>
<td>2.8%</td>
<td>72.7% of deaths are to people aged 65+</td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>7.3%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Numeric Example**

- Population with 100 people
- 10% die at age 1
  - ~ the 1900 infant mortality rate
- If they survive, they live to age 75
- Life expectancy = 0.1(1) + 0.9(75) = 67.6
- Suppose infant mortality rates drops to 1%
  - ~ the 1980 Infant mortality rate
- Life expectancy = 0.01(1) + 0.99(75) = 74.3
Describing determinants of mortality in a cross section

• 323 million people
• ~3.9 millions births
• ~2.7 million deaths

Leading Causes of Death, 2016
• Heart disease 633,842
• Cancer 595,930
• Accidents 146,571
• Chronic lower resp. disease 155,041
• Stroke 140,323
• Alzheimer's 110,561
• Diabetes 79,535
• Influenza/Pneumonia 57,062
• Nephritis 49,959
• Suicide 44,193
### Actual Causes of Death

<table>
<thead>
<tr>
<th>Cause of death</th>
<th># (% of deaths) 1990</th>
<th># (% of deaths) 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco</td>
<td>400,000 (19%)</td>
<td>435,000 (18%)</td>
</tr>
<tr>
<td>Diet/inactivity</td>
<td>300,000 (15%)</td>
<td>400,000 (17%)</td>
</tr>
<tr>
<td>Alcohol</td>
<td>100,000 (5%)</td>
<td>85,000 (5%)</td>
</tr>
<tr>
<td>Micorbial agents</td>
<td>90,000 (4%)</td>
<td>75,000 (4%)</td>
</tr>
<tr>
<td>Toxic agents</td>
<td>60,000 (3%)</td>
<td>66,000 (3%)</td>
</tr>
<tr>
<td>Motor Vehicles</td>
<td>25,000 (1%)</td>
<td>43,000 (2%)</td>
</tr>
<tr>
<td>Firearms</td>
<td>35,000 (2%)</td>
<td>29,000 (1%)</td>
</tr>
<tr>
<td>Sexual Behavior</td>
<td>30,000 (1%)</td>
<td>20,000 (&lt;1%)</td>
</tr>
<tr>
<td>Illegal drugs</td>
<td>20,000 (&lt;1%)</td>
<td>17,000 (&lt;1%)</td>
</tr>
<tr>
<td>Total</td>
<td>1,060,000 (50%)</td>
<td>1,060,000 (48%)</td>
</tr>
</tbody>
</table>

### % that Died in Next 5 Years, Adults, 40-64 Years of Age, NLMS (late 1970)

- **By sex**
  - Males 6.9%
  - Females 3.6%
- **By marital status**
  - Not married 7.0%
  - Married 4.6%
- **By race**
  - Black 7.1%
  - White 4.9%
- **By education**
  - < HS 6.9%
  - HS 4.4%
- **By ethnicity**
  - Non-hispanic 5.2%
  - Hispanic 4.2%
- **By marital status**
  - College 3.6%
- **By income**
  - < $25K 6.0%
  - $25-$50K 3.4%
  - > $50K 2.7%

### Gompertz Equation

- 1825 British actuary Benjamin Gompertz
- "the number of living corresponding to ages increasing in arithmetical progression, decreased in geometrical progression."
- geometrical decrease in survival with age existed because of a geometric increase in the "force of mortality"

\[
M_a = c e^{ba}
\]

- \( M_a \) = mortality rate at age \( a \)
- \( a \) = age
- \( c \) = initial mortality rate
- \( b \) = Gompertz parameter – exponential rate of change in mortality with age

- Note that if \( y = e^{cx} \)
- Then \( \ln(y) = bx \)
- And then \( \ln(M_a) = \ln(c) + ba \)
- Log mortality rates are linear in age
\[ d\ln(M)/da = b \]
\[ d\ln(M) = dM/M = \text{percentage change in } M \]
\[ d\ln(M)/da = \% \text{ change in } M \text{ for a one year increase in age} \]

In the model above:
- \[ \ln(c) = -7.75 \]
- \[ b = 0.0816 \]
- Mortality increases by 8.2\% per year of age

**SES/Health Relationship**

- Health \( (H) \) improves with Socioeconomic status \( (I) \)
- But at a decreasing rate
  - \[ dH/dI > 0 \]
  - \[ dH/dI^2 < 0 \]
- Relationship is true for
  - Nearly all measures of health
  - Nearly all measures of SES (income, wealth, education, status)
  - For all subgroups (by sex, race, age, etc)
  - For nearly all populations
  - For nearly all time periods
  - For nearly all countries
- Focus on one measure of SES = Income
Chetty et al., JAMA 2014

- Match taxpayers (income) aged 40-76 from 1999-2014 to SS death records (mortality)
- 1.4 billion person records
- Income – pre-tax household earnings
  - If file taxes, get from 1040
  - If don't file taxes, get from W2/1099-G (Unemp. comp.)
  - If neither – assume income is zero

Matching income to mortality

- Most people start to collect SS at age 63
- Earnings after this age not a good reflection of their SES status
- If under 63, earnings are the 2 years prior
- If 63 or over, earnings are at age 61
  - Data starts at age 40, years 1999-2014
  - Can follow a 61 year old for an additional 15 years – follow until people are 76
Life expectancy

- Mortality is hard to think about as an outcome
- Expected life expectancy
- If die before age 76 – have actual outcome
- Use Gompertz curves to estimate expected mortality after age 76
- Translate expected mortality into expected lifespan

Difference in life expectancy between top and bottom 1%
Men: 87.3 – 72.7 = 14.6
Women: 88.9 – 78.8 = 10.1

Difference between top and bottom quartile:
13 years growth. Growth rates are .2 and .08 per year. Difference is 13(.2-.08)=1.56 years
Difference between top and bottom quartile: 13 years growth. Growth rates are .23 and .10 per year. Difference is 13(.23-.10)=1.69 years.
Percent Died within 5 years of Survey, Females NLMS

<table>
<thead>
<tr>
<th>Education Group</th>
<th>35-54 years of age</th>
<th>55-64 years of age</th>
<th>65-74 years of age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than high school</td>
<td>2.0</td>
<td>6.0</td>
<td>11.7</td>
</tr>
<tr>
<td>High school graduate</td>
<td>1.3</td>
<td>4.3</td>
<td>9.7</td>
</tr>
<tr>
<td>College graduate</td>
<td>0.9</td>
<td>4.0</td>
<td>8.0</td>
</tr>
</tbody>
</table>

18-64 year olds, BRFSS 2005-2009 (% answering yes)

<table>
<thead>
<tr>
<th>Educ Level</th>
<th>Fair or poor health</th>
<th>No exer. in past 30 days</th>
<th>Current smoker</th>
<th>Obese</th>
<th>Any bad mental hlth past 30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;12 Years</td>
<td>40.9</td>
<td>45.8</td>
<td>37.8</td>
<td>43.6</td>
<td>43.7</td>
</tr>
<tr>
<td>12-15 years</td>
<td>17.8</td>
<td>27.3</td>
<td>26.5</td>
<td>34.7</td>
<td>38.4</td>
</tr>
<tr>
<td>16+ Years</td>
<td>7.2</td>
<td>13.5</td>
<td>10.8</td>
<td>24.8</td>
<td>34.2</td>
</tr>
</tbody>
</table>
Questions for class

• What are the possible mechanisms through which income (or education) can improve health?

• What data supports or refutes each of these hypotheses?
  – List possible explanations
  – Give some evidence for and against
  – Decide whether the pathway is a causal mechanism

What do we mean by causal pathway?

• If causal, we assume that health is determined by income
  – For example, \( H = f(\text{Income}) \)
  
• Therefore, \( \frac{dH}{dI} > 0 \)
  – An exogenous change in income will alter health

• Example: Suppose we change social security benefits – if income is causal, this should alter mortality of the elderly

Why is it hard to determine whether the income/health relationship is causal

• Many factors that determine high income
  – Drive/ambition/intelligence/risk taking/luck/background

• Many of these same factors can also impact health

• Therefore, we do not know whether income is causing better health, or some third factor that is unmeasured

\[
\text{Died}_i = \alpha + x_i \delta + \text{income}_i \beta + \epsilon_i
\]

\( \text{Died}_i = 1 \text{ if died within 5 years, } 0 \text{ otherwise} \)

\( x_i = \text{controls} \)

\( \text{income}_i = \text{annual family income} \)

\( \hat{\beta} \text{ unbiased is } E(\epsilon_i | x_i, \text{income}_i) = 0 \)
Problem:

- Realization of $\varepsilon_i$ conveys information about income
- If $\varepsilon_i > 0$, more likely to die early
- Could mean you had lower income because you were sick and could not work as much (reverse causality)
- Could mean you have a high discount rate – don’t invest in human capital for the job market (which means lower income) and it means you maybe did not invest in health (which means higher $\varepsilon_i$)

Story we are telling is that $\text{cov}(\varepsilon_i, \text{income}_i) < 0$

- We believe $\beta < 0$
- This means we are “overstating” the impact of income on mortality –

Table 3

<table>
<thead>
<tr>
<th>Economic Effects of New Health Onset</th>
<th>Wealth</th>
<th>OOP Expenses</th>
<th>Total Medical Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild onset</td>
<td>$-3,620$</td>
<td>$655$</td>
<td>$2,565$</td>
</tr>
<tr>
<td>Severe onset</td>
<td>$16,846$</td>
<td>$2,565$</td>
<td>$28,563$</td>
</tr>
<tr>
<td>AHEAD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any onset</td>
<td>$-10,451$</td>
<td>$1,056$</td>
<td>NA</td>
</tr>
<tr>
<td>HRS severe onset only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With health insurance</td>
<td>$-17,417$</td>
<td>$1,912$</td>
<td>$29,957$</td>
</tr>
<tr>
<td>Without health insurance</td>
<td>$-17,282$</td>
<td>$4,575$</td>
<td>$42,166$</td>
</tr>
<tr>
<td>HRS severe onset only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below median income</td>
<td>$-11,346$</td>
<td>$2,493$</td>
<td>$29,893$</td>
</tr>
<tr>
<td>Above median income</td>
<td>$-25,371$</td>
<td>$2,614$</td>
<td>$28,983$</td>
</tr>
<tr>
<td>AHEAD any onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below median income</td>
<td>$-4,977$</td>
<td>$913$</td>
<td>NA</td>
</tr>
<tr>
<td>Above median income</td>
<td>$-17,040$</td>
<td>$1,185$</td>
<td>NA</td>
</tr>
</tbody>
</table>
Clark and Royer

- Examines education/health link using shock to education in England
- 1944 law
  - Raised age of comp. schooling from 14-15
  - Went into effect April 1, 1947
  - Raised comp. years of schooling to 9
  - Gave Minister of Ed power to increase to 16 under certain conditions
  - Did so in Sept 1, 1972
    - Raised comp. years of schooling to 10

- Produce large changes in education across birth cohorts
- Changes in education and health are “smooth” over birth cohorts
- If education alters health, should see a structural change in outcomes across cohorts as well
- What assumptions have to be true for this to generate an unbiased estimate of the impact of schooling on health?

Figure 1: Years of full-time education by quarter of birth

Notes: Data for those not enrolled in grades 1991-1994 at the Bohdi Survey of England. The revised law was enacted corresponding to the first cohort subject to the new compulsory schooling law. The first of these took effect on 1 April 1947, the second on 1 September 1972.
This figure is not in the paper but in a previous version. It shows birth cohorts versus $\ln$ (gross weekly earnings). What does this graph show and why is this informative?

Sullivan and von Wachter

- Consider the opposite of Gardner and Oswald – what happens when someone loses income
- Lost income due to job loss
- Focus on displacement?
  - What is displacement?
  - Why displacement and not job loss?
Data

- 5% random sample of unemployment records in PA 1974-1991
- Have quarterly earnings
- Select sample of workers with the same employer 1974-1979 (firms > 50 workers)
- Identify people who have been “displaced”
  - Lose job 1980-1986
  - And when firm size falls by 30% or more

Impact of displacement on earnings

\[ y_{it} = \alpha_i + \lambda_t + x_{it}\beta + \sum_{k=-20}^{19} D_k^i \delta_k + \epsilon_{it} \]

\( i = \text{person, } t = \text{quarter} \)
\( y_{it} = \ln(\text{quarterly earnings}) \)
\( \alpha_i = \text{person effect} \)
\( \lambda_t = \text{quarter (time) effect} \)
\( x_{it} = \text{time-varying characteristics} \)
\( D_k^i = 1 \text{ if person } i \text{ was displaced } k \text{ quarters ago (after)} \)
\( \delta_k = \text{effect of displacement} \)

Impact of displacement on mortality

\[ y_{it} = \alpha_i + \lambda_t + x_{it}\beta + D_t^i \delta + \epsilon_{it} \]

\( i = \text{person, } t = \text{year} \)
\( y_{it} = 1 \text{ if person dies in period } t, 0 \text{ otherwise} \)
\( D_t^i = 1 \text{ if person } i \text{ was displaced in the year} \)
\( \delta_k = \text{effect of displacement} \)
7.1/1000=0.007 mortality risk among non-displaced workers
Stress as an explanation for the SES/Health Gradient

- Usual suspects don’t explain gradient
- Leading candidate is Stress
- Low SES face more persistent stress
- Body reacts to stress in a good way in the short run
- Persistent stress can cause more permanent damage

HPA Axis

- Hypothalamic-pituitary-adrenal axis
- Put into work when the body faces stress
- Regulates many body functions including digestion, immune, mood, emotions, energy storage
- Concern: activation of system is “good” under stress, but it does come at a cost. Therefore, persistent stress generates more permanent damage to the body’s systems

Cortisol

- Circadian rhythm. Rises when awake, in late afternoon
- Regulates many activities
- Under stress, more cortisol is produced
  - Increases availability of glucose
  - Suppresses energy available to other systems like immune
  - Cortisol reduces after the stress subsides
- Problems
  - Constant stress leads to dysregulation of HPA
  - Stress in early life can generate dysfunction of HPA
**Cortisol**

- Stress increases cortisol
  - Higher among residents
  - Higher among accountants near April 15th

- Poor have elevated cortisol at all times
  - They are more exposed to stress

- Elevated cortisol thought to
  - "burn out" major organs – they just work harder
  - Increases susceptibility of immune system

---

**Primate research**

- Observational studies show worse health among subordinate male baboons
  - Elevated stress hormone (glucocorticoid) levels, worse cholesterol profile

- Experimental manipulation of status provides more compelling evidence
  - Causal effects of subordination and harmful effects of "status competition"

---

**Baseball Hall of Fame**

- **Baseball Writers Association of America**
  - Annual voting held since 1936
  - Eligibility: >10 seasons in MLB, retired 5+ years, max of 15 ballot appearances
  - Voting: ~450 voters, mail-in ballot, can name up to 10 players
  - Induction: Must be named on 75% of total ballots cast
  - Compete voting results are reported to public (newspapers)

- **Committee on Baseball Veterans (Veterans)**
  - Select former MLB players not chosen by BBWAA
  - Historically voting was held annually
  - Much smaller committee (~15), but similar 75% required for induction
  - Voting results not publicly disclosed and accusations of cronyism
  - Major reforms in 2001 (expanded voting pool, public disclosure)
Sample

- All players alive while appearing on at least one ballot between 1945-2006
- Restrict analysis to pre-1946 births to reduce censoring (N=597)
- Key derived variables:
  - Indicators of induction status (BBWAA and veterans)
  - Maximum vote share ever received (categorical: <1, 1-2, ..., 51-74, 75-78, ...)
  - Number of “close losses” (defined as vote share $\geq 50$ but $<75$)
  
Adjusted life duration by maximum vote share

Cause of death by maximum vote share