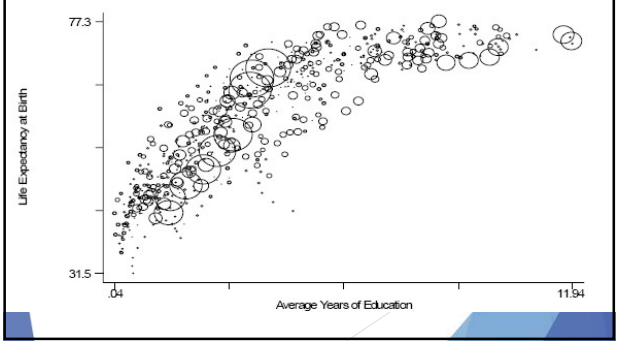
Regression analysis in the context of health/labour economics

- The relationship between education and health:
- $H_i = \alpha + \boldsymbol{\beta} E_i + \boldsymbol{X}_i \delta + \varepsilon_i$
  - where H_i is a measure of *i*'s health (or health behavior), E_i is education, X_i is characteristics like race, gender and age dummies.
- The coefficient on education **β** (referred to as the "**education gradient**") is the object of interest.

Health and Education

- Empirically there seems to be a **very strong and robust correlation between education and health**
  - More highly correlated than income even
    - Income effect seems to get smaller
  - Effect of schooling seems to remain strong even at high levels of education
- Often referred to as "**health gradients**"

Barro-Lee international data



Policy implications

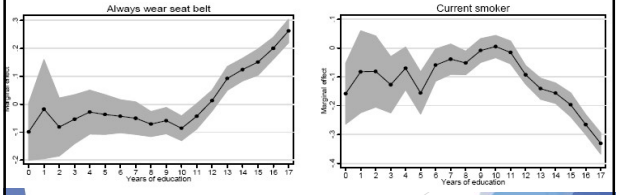
- Substantial attention has been paid to these "**health inequalities**."
- Gradients in health by education are now being systematically monitored in many countries
  - the US included them as part of its Healthy People 2020 goals
- Some countries (eg UK and Australia) have explicit target goals of reducing health disparities associated with inequality

Health conditions

- The slope of the gradient between education and health varies across conditions
  - And they are not trivial in size- they are quantitatively important
- Cancer is the exception
  - few studies find that education is associated with lower cancer mortality.

Why are E and H are correlated?

- Better health behaviours
  - But this accounts for only some of the effect of E on H



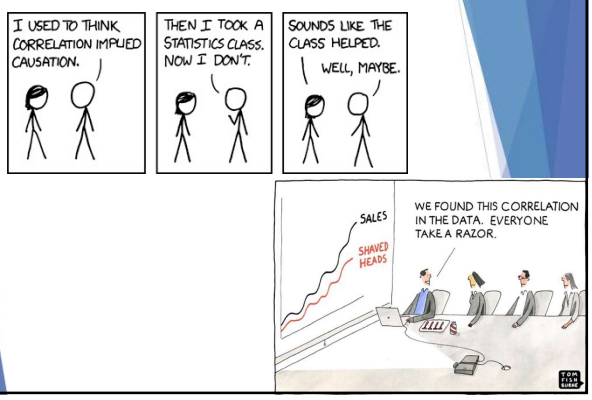
Heterogeneity in effects

- Note that these effects are approximately linear (beyond 10 years of education)
- Effects differ across **gender**
  - depression and obesity gradient bigger for women
  - Mortality and dinking gradients bigger for men
- Some gradients tend to fall with **age**
  - Seat belt use, smoking
- Other gradients increase with age
  - Functional limitations, depression

Heterogeneity in effects

- Effect of E on H usually found to be smaller for poor than non-poor
  - **E and income may be complementary inputs** in production
    - E gives people better information, income allows them to act on it?
  - Health for the educated getting better faster
- Education seems to have **spillover effects**
  - Educating mums improves children health

Correlation and Causation



Is the effect of E on H causal?

- Three reasons for link between E and H:
  - Poor H leads to low E
  - E improves H
  - E and H are both driven by something else
- Important to know extent to which each is true- **policy implication**:
  - E subsidies only improve H if E ***causes*** H

Causality?

- Children with poor H may obtain less E
  - Low birth weight associated with childhood poor H and hence low E
  - and low birth weight associated with later poor H

Causality?

- **Unobservables** (genes, family background, preferences, etc) may affect E and H
  - Richer parents may invest in both more H and more E for their children
  - Higher preference for future outcomes imply high E and higher H
- But attempts to control for such factors do not explain **all** of the effect of E on H
  - E.g. by adding family background controls
  - E generally remains large and significant

Methodology

- Cross sectional results cannot be interpreted as **causal**
- Some problematic issues are:
  - **Endogeneity** (which is the direction of causality? Is it from labour market to health or *viceversa*?)
  - **Measurement error** (people may not report circumstances correctly, especially their health conditions)

Methodology

- Other problematic issues are:
  - **Omitted variable bias** (the real effect comes from variables we do not include in the regression)
  - **Unobserved Heterogeneity** (people have unobserved characteristics, i.e. ability, personality, temper which may affect both their health and their labour market outcomes)
  - **Right /Left censoring** (we do not observe the entire life of a person. We often start observing after initial conditions have taken place or loose individuals without knowing the end of their story!)

The use of longitudinal data

- **Longitudinal data involve repeated observations of the same individual over long periods of time**
- Longitudinal data are different from cross-sectional that involve observation of all of a population, or a representative subset, at a defined time
- This is the type of data we'll use in this class!

The use of longitudinal data

ADVANTAGES:

- Invaluable information about **changes in the lives of people**
- Keep into consideration **CONSTANT** characteristics
- Look at the effects of earlier life circumstances on later outcomes in a person's life
- The large size: investigate differences between regional areas