Education and coronary heart disease risk associations may be affected: By early-life common prior causes: A propensity matching analysis

Loucks EB, Buka SL, Rogers MI, Liu T. (Brown University)
Kawachi I, Kubzansky LD, Gilman SE. (Harvard School of Public Health)
Martin LT. (RAND corporation)


Background

Education is inversely associated with coronary heart disease (CHD) and it’s risk factors (smoking, BP, obesity and diabetes)

However, the causality remains unknown in presence of childhood intelligence, childhood socioeconomic circumstances, parental mental illness, childhood chronic health conditions, birth weight, among others.

AIM I

Traditional confounders (age, sex, and race/ethnicity)

Education is associated with lower 10-year risk for CHD (by using the validated Framingham risk algorithm), independent of a wide range of early childhood conditions.

AIM II

Traditional confounders (age, sex, and race/ethnicity) + 21 early life potential confounders

Associations of educational attainment with individual modifiable CHD risk factor components of the Framingham algorithm to identify if any aspects of CHD risk are particularly strongly associated with education.

Sample

17,921 offspring from New England Family Study (NEFS) under collaborative Perinatal Project (CPP) between 1959 and 1974

914 participants were selected (preference for racial/ethnic minorities, low or high educational attainment) from EdHealth Study (formerly NEFS) between 2005 and 2007

38–47 years

16 excluded (not living, incarcerated)

280 didn’t participate

42 participants were not interviewed in person and did not complete physiological assessments

5 reported angina or myocardial infarction

18 participants missing >3 Framingham risk algorithm components

168 eligible participants

618 participated

Final participation n = 553
Primary exposure variable

EDUCATION (Self-reported)

- ≤ high school (< high school, high school degree, or GED)
- Post secondary training (schooling after high school that did not result in a bachelor’s degree or greater)
- college degree (e.g., bachelor’s degree, graduate degree)

Outcome variable

- Smoking (self reported) – Y/N
- Lipids – measured in non-fasting plasma samples
- Total and HDL cholesterol – measured enzymatically
- Presence of diabetes (self-reported) – Y/N
- Five systolic and diastolic blood pressure were obtained over 1-minute intervals, the mean of the lowest three systolic or diastolic blood pressure readings was recorded, excluding the first recorded blood pressure.

Outcome variable

10-year risk of CHD

- Percentage (31% chance of developing CHD over 10 years)

Framingham risk algorithm

10-year risk of CHD

- Algorithm uses sex-specific age-adjusted Cox proportional hazard (CPH) regression models that incorporate diabetes, total and HDL cholesterol, systolic and diastolic blood pressure, smoking, age and sex
- The 12-year follow-up was used in the CPH models, and results were adapted to provide 10-year CHD incidence estimates.
- Separate score sheets were developed for each sex using TC and LDL-C categories.
- Score sheets for prediction of CHD using TC and LDL-C categorical variables were developed from the $\beta$-coefficients of CPH models.

Childhood determinants of educational attainment

- Parental socio-economic position – weighted percentile of both parents’ educational attainment, occupation, and income relative to the U.S. population prenatally and at age 7
- Father’s absence from household at birth or age 7 (Y/N)
- Household crowding (>1.5 persons/room) prenatally and age 7
- number of moves between birth and age 7
- Parental demographic factors included age at birth, and marital status at birth and age 7
- Mothers’ and fathers’ treatment for psychiatric and substance use disorders were obtained by mother’s self-report during pregnancy and at offspring age 7

Source: Prenatal to age 7 years from CPP
Continued....

- Childhood health status was assessed based on exposure to
  - maternal smoking during pregnancy;
  - birth weight (g), and
  - chronic medical conditions identified between birth and age 7
- Childhood chronic medical conditions were derived from
  physical examinations by CPP pediatricians at ages 1 and 7 years
- Summary score of number of chronic physical health conditions
  (including abnormalities of the liver, cardiovascular conditions,
  hematologic conditions, lower respiratory tract abnormality,
  neoplastic disease, neurologic abnormality, and prolonged
  /recurrent hospitalization)
- Childhood intellectual development at age 7

Source: Prenatal to age 7 years from CPP

RCT vs propensity score matching

Conceptual similarity

Both study designs attempt to overcome problems of confounding bias by obtaining samples that differ only with respect to educational attainment

Propensity score (PS)

- Gather all the background information that the participants have before treatments are assigned, that might plausibly affect which treatment they get
- Build a model to predict the probability that they will receive the treatment instead of the control.

\[ PS = \Pr(\text{education group} \mid \text{background information}) \]

- Groups of subjects with similar propensity scores can then be expected to have similar values of all of the background information, in the aggregate.

In propensity score matching,

- More educated individuals are matched to a sample of less educated individuals who are similar with respect to the set of childhood background characteristics.
- Used “caliper matching” to construct the matched samples, in which participants are randomly sorted, and then each higher-educated individual is matched to the lower-educated individual with the closest propensity score, as long as their propensity scores are within 10 percentage points of one another (ie, caliper width of 0.10)
- Higher-educated participants who were unable to be matched to a lower educated control according to these criteria were excluded from analyses.

PS = \Pr(\text{education group} \mid \text{background information})

- The conditional probability of receiving a given exposure (Education) given a vector of measured covariates.
- Usually estimated using logistic egression
- Two models:
  - some post-secondary training (1) vs. ≤ high school (0) and
  - college degree (1) vs. ≤ high school (0)
- Predictors all of the childhood determinants of educational attainment
The first model included only age, sex, and race/ethnicity.

The second model included age, sex, race/ethnicity, and the childhood determinants of education.

The 10-year risk of CHD was analysed by the use of linear regression.

The distribution of the 10-year CHD risk variable was skewed, and consequently log (natural) transformed.

Use [(exp(β) - 1) × 100] to maintain the original units of the CHD risk algorithm

Analyses used generalized estimating equations to account for clustering by family
Strengths

- Accurate measurement of biological measures by the use of substantial internal and external quality control protocols.
- The birth cohort study design

Conclusion

Participants with a college degree had lower CHD risk (27.9% lower CHD risk) after accounting for traditional confounders (age, sex, and race/ethnicity).

Further statistical accounting for early life potential confounders (such as intelligence, childhood economic circumstances, childhood chronic illness, parental mental health) resulted in a moderate effect size (13.1% lower risk).

A reasonable amount of the observed educational disparities in CHD risk found in observational studies may be due to factors occurring earlier in the life course.

What I didn’t understand!

- Multiple imputation (p224, left column, 2para)
- Why standardised difference? $d = \frac{100 \times (\bar{x}_2 - \bar{x}_1)}{\sqrt{\frac{s^2_{\bar{x}_2} + s^2_{\bar{x}_1}}{2}}}$
- Sensitivity analysis for different caliper (0.01-0.20) has been used in the matching procedure, but how they applied the sensitivity results in the multiple imputation?
- Due to low sample size and large number of adjustment variables, smoking analyses were performed using logistic regression on pairs discordant for smoking; point estimates and 95% CI are odds ratios (last footnote - Table 3)!
- ..........