Published in final edited form as:
Am J Epidemiol. 2013 May 1; 177(9): 882–884. doi:10.1093/aje/kwt042.

Invited Commentary: Can Changes in the Distributions of and Associations Between Education and Income Bias Estimates of Temporal Trends in Health Disparities?

Makram Talih, Ph.D.
Office of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention

Abstract

Chen et al (*Am J Epidemiol* 2012; xx:xxx—xxx) develop a simulation study for comparing various measures of socioeconomic health disparities when bias can arise from temporal changes in the bivariate distribution of education and income. In this commentary, I argue that, in relation to health, the "meaning" of education cannot be reduced to its socioeconomic value; improved health literacy, for instance, can result in important health benefits. Further, I suggest that unless there is a substantial prior understanding of the data generating mechanism, directed acyclic graph models should be avoided because causal relationships cannot be inferred from regression. An alternative is to resort to conditional independence graphs, which use only undirected edges. Finally, although the slope index of inequality (SII) can, in some specific cases, be seen to reduce bias in temporal comparisons of socioeconomic health disparities, it was not designed for causal inference. The SII simply describes the average change in the proportion in poor health when the population is ordered by socioeconomic status.

Keywords

health literacy; conditional independence; undirected graphs

Numerous indicators of health have shown improvement in the United States (U.S.) in recent decades. One of four overarching goals of Healthy People 2020 remains to "achieve health equity, eliminate disparities, and improve the health of all groups" (1), yet health disparities persist between populations (2). The Centers for Disease Control and Prevention (CDC) defines social determinants of health as the complex, integrated, and overlapping social structures and economic systems (e.g., the social environment, physical environment, health services, and structural and societal factors) that are responsible for most health inequities (3). Social determinants of health are shaped by the distribution of money, power, and resources throughout communities, and closing the health gap that results solely from

---

Correspondence to: Makram Talih, Ph.D., National Center for Health Statistics, 3311 Toledo Road, Room 6317, Hyattsville, MD 20782 (mtalih@cdc.gov).

Disclaimer: The opinions in this commentary are those of the author and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Conflict of Interest: None

one's socioeconomic position (SEP) is both a national (4) as well as an international priority (5).

SEP is a multidimensional construct that includes wealth, income, education, and occupation, which are all linked to social class (3,6). Higher income can benefit health through improved access to material resources such as higher quality diets, shelter, and health care. Higher income can also result in a higher status and power within one's community, further reinforcing access opportunities (7). Of course, the components of SEP are interrelated. Chen et al (6) focus on the relationship between education and income. For instance, higher parental income can provide offspring with a better quality education, and higher educational attainment, in turn, serves as credential for employment in more prestigious and greater income-generating occupations.

Chen et al (6) argue that the value or "meaning" of educational attainment has changed over time; for example, they observe that returns on a high school degree in the form of income have diminished. This observation serves as the impetus for their simulation study, which aims to examine how changes in the bivariate distribution of income and education can bias findings of health disparities. While this is a compelling argument, and while it certainly warrants the simulation study that Chen et al (6) report on, my opinion is that the argument is reductive; education is equated only with its socioeconomic value.

Education plays a key role in influencing the behavior of consumers of the health care system. More education empowers individuals to better assess risks associated with their behavioral and life style choices. Individuals can also be more effective advocates for improving conditions in their work and living environments. They are more likely to adhere to treatments, follow prescriptions, and understand labels. The capacity to understand basic health information and make appropriate health decisions, i.e., health literacy, plays a major role in the delivery of high-quality care and is at the forefront of Federal health policy. Federal policy initiatives such as the National Action Plan to Improve Health Literacy and the 2010 Plain Writing Act aim to boost health literacy and help move the nation beyond the cycle of costly 'crisis care' (8).

Changes to the bivariate distribution of income and education can occur either in the marginal distribution of income or education or in the covariance between income and education. Changes can also occur in how those two components of SEP predict a health measure such as body mass index (BMI). However, Chen et al (6) do not articulate clearly why causal modeling is appropriate in the context of examining the (time-dependent) relationship between socioeconomic position and BMI. Unless there is a substantial prior understanding of the data generating mechanism, causal relationships cannot be inferred from regression; only associational relationships can be inferred (9). Yet, on the one hand, the causal relationship between income and education is not well understood, because the labor market is complex and because SEP determines and is determined by one's career trajectory through the labor market (10). On the other hand, in relation to health, even though Chen et al (6) and others argue that education is less prone to reverse causation than income—in general, schooling is completed before individuals develop the chronic health conditions of adulthood—chronic health conditions in childhood can impact an individual's educational attainment (7). Moreover, added income can be detrimental to health outcomes —i.e., a causal relationship can be refuted—as shown for BMI in recipients of the Earned Income Tax Credit in the U.S. (11).

In light of these concerns, and because the principal objective of the simulation in Chen et al (6) is to examine the bias in findings of socioeconomic health disparities that arises from temporal changes in the joint distribution of income and education, the authors could forego the directed acyclic graph (DAG) and simply resort to a conditional independence model, which uses only undirected edges (12). Treating all data as continuous, this would be equivalent to restricting some off-diagonal elements to zero in the inverse of a 6×6 variance-covariance matrix in a multivariate normal distribution—using the notation in Chen et al (6), the multivariate random vector (X, Y, Z), at two time points, is 6-dimensional. Doing so would avoid lingering questions about the causal relationship between income and education and about the etiology of the socioeconomic gradient in BMI. In addition, such a fully multivariate approach would enable significance tests of the difference in observed estimates between time periods that adjust for correlation between time points.

Another source of concern is with treating educational attainment as a continuous variable in the DAG of Chen et al (6) or in the conditional independence model just outlined. Educational attainment is best treated as a discrete variable because its values tend to cluster according to the major educational milestones in one's lifespan (high school degree, Associate's degree, college degree, etc.) Further, it can be argued that not every year of additional education contributes equally to better health (7), which violates the linearity assumption in Chen et al (6).

Finally, Chen et al (6) use the simulation for a comparative study of various measures of socioeconomic health disparities. In particular, the authors assert that the slope index of inequality (SII) was developed explicitly to minimize bias in temporal comparisons. In my view, the SII is no more than a descriptive statistic, interpreted as the average change in the proportion in poor health (e.g., adults aged 20 and over with BMI ≥30) when the population is rank-ordered by income or education. Said another way, SII is similar to looking at the excess adverse health (e.g., obesity) in the lowest ranked group compared with the highest ranked group (13). In practice, the SII is especially useful when continuous individual-level data are unavailable for analysis, but only the corresponding group-level data are available. For example, both Healthy People 2020 (1) and Healthy People 2010 (2) use population templates where education and income categories are dictated by U.S. Office of Management and Budget (OMB) standards.

The analysis that Chen et al (6) have undertaken is an important one because it sheds light on bias that can arise when comparing socioeconomic gradients in health over time. That such bias can occur in even the simplest of models, as the authors show, raises serious concerns about the extant findings of socioeconomic health disparities that do not adjust for changes in the distribution of and relationship between education and income. I agree with the authors that understanding how the joint distribution of education and income changes over time is essential to a meaningful analysis of socioeconomic health disparities. Yet, in relation to health, I caution that education cannot be reduced to its socioeconomic value;

improved health literacy, for instance, can result in important health benefits. Further, I suggest that unless there is a substantial prior understanding of the data generating mechanism, directed acyclic graph models should be avoided because causal relationships cannot be inferred from regression. An alternative is to resort to conditional independence graphs, which use only undirected edges. Finally, although the slope index of inequality (SII) can, in some specific cases, be seen to reduce bias in temporal comparisons of socioeconomic health disparities, it was not designed for causal inference. The SII simply describes the average change in the proportion in poor health when the population is ordered by socioeconomic status. Nonetheless, I agree with the authors that analyses that rely on the SII should, to the extent possible, control for changes in the joint distribution of education and income.

List of Abbreviations (Main Text)

U.S	United States
CDC	Centers for Disease Control and Prevention
SEP	socioeconomic position
BMI	body mass index
DAG	directed acyclic graph
SII	slope index of inequality
OMB	Office of Management and Budget

References

1. Department of Health and Human Services. Healthy People 2020. Washington, D.C: U.S. Department of Health and Human Services; 2010. (http://healthypeople.gov/) [Accessed October 30, 2012]
2. National Center for Health Statistics. Healthy People 2010 Final Review. Hyattsville, MD: National Center for Health Statistics; 2011.
3. Centers for Disease Control and Prevention. Social determinants of health: definitions. Atlanta, GA: U.S. Centers for Disease Control and Prevention; 2011. (http://www.cdc.gov/socialdeterminants/Definitions.html) [Accessed October 30, 2012]
4. Institute of Medicine. For the public's health: the role of measurement in action and accountability. Washington, D.C: Institute of Medicine; 2010.
5. World Health Organization. Closing the gap in a generation: health equity through action on the social determinants of health—Final report of the Commission on Social Determinants of Health. Geneva: World Health Organization; 2008.
6. Chen JT, Beckfield J, Waterman PD, Krieger N. Can changes in the distributions of and associations between education and income bias estimates of temporal trends in health disparities?—An exploration with causal graphs and simulation. Am J Epidemiol. 2012; xx:xxx–xxx.
7. Kawachi I, Adler NE, Dow WH. Money, schooling, and health: mechanisms and causal evidence. Ann N Y Acad Sci. 2010; 1186:56–68. [PubMed: 20201868]
8. Koh HK, Berwick DM, Clancy CM, et al. New Federal policy initiatives to boost health literacy can help the nation move beyond the cycle of costly 'crisis care'. Health Aff. 2012; 31:434–443.
9. Freedman DA. Graphical models for causation, and the identification problem. Eval Rev. 2004; 28:267–293. [PubMed: 15245621]
10. Scott MA. Affinity models for career sequences. J R Stat Soc (Series C). 2011; 60:417–436.

11. Schmeiser MD. Expanding wallets and waistlines: the impact of family income on the BMI of women and men eligible for the Earned Income Tax Credit. Health Econ. 2009; 18:1277–1294. [PubMed: 19142860]
12. Whittaker, J. Graphical models in applied multivariate statistics. Chichester, UK: Wiley; 1989.
13. Keppel K, Pamuk E, Lynch J, et al. Methodological issues in measuring health disparities. National Center for Health Statistics. Vital Health Stat. 2005; 2(141)