An economic perspective on the causal explanations for the socioeconomic inequalities in health

Lori J. Curtis

ABSTRACT

Socioeconomic inequality, or the socioeconomic status (SES) gradient, is arguably one of the most-studied phenomena in health. The gradient in health is apparent in objective and subjective measures, across virtually all countries, and is evident at individual and population levels. There is no longer much debate over the relationship between SES and health. However, exact causal pathways remain elusive. Advocating for strong policy to reduce or eliminate the SES-health gradient necessitates understanding the causal pathways, from intervention to outcome. While economists are not convinced that there is a clear enough understanding of the causal pathways of the SES-health gradient, they have produced a substantial body of work from which to move forward. The article briefly discusses the theoretical underpinnings used by economists as a basis for the study of the causal pathways for the health gradient. That presentation is followed by a concise overview of some of the evidence that economists have produced. The paper concludes with a discussion of how current economic evidence may be used to help policymakers advocate for interventions to limit the SES gradient in noncommunicable diseases.

Keywords Health inequalities; economics.

Opinion and analysis Curtis • Causal explanations for socioeconomic inequalities in health

The difficulty in demonstrating causal links intensifies because reverse causality and endogeneity are particular problems in the study of the SES-health relationship (5). As previously discussed (7, 8), the primary premise is that higher SES leads to better health. However, it is possible that the causal path runs from health to SES. For example, lower health status could restrict education, human capital accumulation, and labor market participation, thus leading to lower education and/or income (reverse causality). Alternatively, unobserved characteristics may influence both education and health decisions, so the effects of one on the other cannot be estimated consistently (5, 11). Economists have attempted to identify unobserved characteristics within socioeconomic strata that may drive health differences (8, 12-15). These characteristics include access to health care (utilization is typically assessed with surveys that have no measure of when care was needed and not received, or the reasons); differential productivity in the use of health information or health care; differential vulnerability; such environmental factors as pollution or chemical exposure (13-15); and cumulative effects (i.e., individuals from lower socioeconomic situations experience more health shocks than those from higher socioeconomic situations) (16-20).

This short paper presents a brief discussion of the theoretical underpinnings used by economists as a basis for the study of the causal pathways for the health gradient. This is followed by an overview of some of the widely cited economic research in this area. The overview is organized by the causal pathways that have been studied: SES → health; SES → health behaviors → health; health → SES (reverse causality); and unemployed factors → SES and health (unobserved heterogeneity). There is next a brief discussion of evidence available in the Americas and for noncommunicable diseases (NCDs) specifically, followed by some conclusions and policy discussion.

The paper is not meant to be an exhaustive or systematic review, but instead to provide an overview of some of the well-cited, highly regarded evidence that economists have produced. The overview covers studies examining secondary data using various methodologies, as well as reviews of such studies. It is anticipated that this presentation of economic evidence will lead readers to investigate further how current methodologies may be used to develop new frameworks and expand the use of economic theories and econometric tools to effectively advocate for the prevention of NCDs, including developing policies within and beyond the health sector in the Americas and other world regions.

THEORETICAL UNDERPINNINGS

Standard economic theory tells us that individuals choose to consume goods and services based on their preferences, budget constraints (e.g., different incomes and prices), and beliefs in the impact of their actions (if the assumption of full information is not invoked).

Most economic theories on the relationship between SES (typically proxied by income and/or education) and health expand the standard theory based on one of the seminal papers written by Becker (21, 22), Grossman (6), and/or Rosenzweig and Schultz (23). The authors add differences in productivity, household production, and unobserved characteristics to the standard framework. Basically, the models maximize utility (happiness, well-being), which is created by consuming goods and services (G&S) and health. The G&S may positively influence health (healthy food, etc.) or negatively do so (cigarettes, etc.). As well as purchasing final goods in the market (e.g., a healthy meal), individuals may use their time to convert (via a production process) goods purchased in the market (intermediate goods (e.g., ingredients to make the healthy meal)) into final consumption goods (the healthy meal) and health. The ease of the production process (productivity) depends on the individual’s level of education and other unobserved factors (e.g., family background, genetics, intelligence). The amount of goods purchased in the market and the time available for production depends on prices, wages, labor market participation, income, and wealth.

Although empirical studies often avoid laying out explicit theoretical models, the general framework underpins much of the research. The theoretical models (6, 21-23) implicitly direct researchers to focus on the pathways between SES, health inputs (e.g., health care, health insurance (if public health insurance is not available to all), health behaviors (positive and negative), and unobserved factors (unobserved heterogeneity (e.g., genetics, intelligence), environment, etc.)), and health outcomes. Expansive literatures have evolved on the SES → health pathways, including the SES → health behaviors → health pathway.

PATHWAYS INVOLVING THE SOCIOECONOMIC STATUS (SES) GRADIENT

SES → health

Excellent overviews on the relationship between SES (income (2) and education (10, 14)) and health are available. Evans, Wolfe, and Adler (2) conclude that “despite much work on the mechanisms that lie behind the gradient (24), we cannot fully account for the observed disparities in health across income.” As previously stated, education has been used as a more robust instrument for SES than income (2). An oft-cited paper by Cutler et al. (13) uses multiple nationally representative cross-sections of survey and administrative data from the United States of America to show that for non-Hispanic whites the educational gradient in mortality (from cancer and cardiovascular disease (CVD)) has grown over time and that this cannot be explained by changes in key behavioral risk factors. The returns to education (conditional on health behaviors) and changes in returns to health behaviors

2

Rev Panam Salud Publica 42, 2018
are important and increasing, and they are stronger for males than females. The impact of smoking strengthened over time for both men and women, as have the consequences of severe obesity for females. Results suggest that complete elimination of disparities in behavioral risks across education groups would likely not decrease the differentials in mortality substantially (approximately 7% to 25%). Although influential, the study uses cross-sectional data. This type of data is strongly criticized in the literature for its inability to identify causality.

The shortcomings of cross-sectional data propelled the use of natural experiments to study causal relationships. Natural experiments occur when circumstances outside the control of the researcher (e.g., the introduction of a public policy such as compulsory schooling reform) lead a subset of a population to be differentially exposed to a hypothesized causal factor (i.e., education). Significant differences in outcomes across exposed and unexposed populations can indicate a causal path between the exposure and the outcome. A substantive review of the schooling reforms literature is offered by Meghir et al. (15). A strong positive relationship between education and health is found in the United States in data from the early to mid-1900s (25). However, the same data produces more muted results when state-specific time trends are included with the policy reforms or when larger datasets (which identify individuals rather than cohorts) are employed (26). Given that results differ when using data from the same country, it is not surprising that conclusions examining compulsory schooling reforms are mixed when using data from different countries and time periods. Positive results are found in Denmark (27) and England and Northern Ireland (28) by some researchers, while others find negligible or negative results in Sweden (15) and the United Kingdom (29). Results are strongly dependent on where and when the data originate (perhaps pointing to differences in institutions across countries, sample sizes, and populations (15)) and the empirical analyses employed (26).

Cross-country comparisons also produced inconsistent results. Using countries with changes in mandatory leaving age as an instrumental variable in a dynamic health equation, Brunello et al. (9) find an additional year of education decreases self-reported poor health by 7% for females and by 3% for males. Health behaviors explain approximately one-quarter of the effect. Moreover, Cutler and Lleras-Muney (6) show that income, health insurance, and family background explain about one-third of the education gradient in health, and that cognitive ability explains about one-fifth of it. They demonstrate a pathway running from education to cognitive ability to healthier behaviors to health. Discounting, risk aversion, or time preferences account for none of the gradients in health behaviors. There is some evidence that the social environment (healthier for the better educated) accounts for about one-tenth of the education/health gradient.

The mixed results found in the literature spurred researchers to identify better ways of identification. Studies using samples of monozygotic (identical) twins to control for unobserved factors (e.g., family and genetic backgrounds shared completely by identical twins) were thought to be a solution. These studies suggest that causal impacts of schooling on health outcomes and behaviors are much smaller than suggested by other studies (although some of the studies were criticized for small sample sizes). Amin, Behrman, and Spector (30) used multiple twin registries in the United States and found, like many other studies, that schooling is significantly associated with numerous health outcomes and behaviors. However, no causal relationship could be identified between schooling and better health behaviors after controlling for unobserved factors. Twin studies have found that higher education is positively related to self-reported health (10, 31) and negatively related to the number of chronic conditions (31), but causality has been more difficult to assign.

A multidisciplinary review of the literature (32), which includes many of the studies discussed in this paper, claims that there is a sufficient body of evidence to suggest that schooling is causally related to improvements in health outcomes, and that raising the incomes of the poor leads to improvement in their health outcomes. However, that review also notes that the findings are crude and that more specific questions need to be asked, such as what type of education matters for health or whether there is a difference between the health impacts of temporary income shocks versus changes in long-term income. Other reviews of the evidence tend to support the need for further evidence.

Health → SES (reverse causality)

There is a plethora of literature on the relationship between childhood health and outcomes in later life. The hypothesis is that the causal pathway runs from child (poor) health to (lower) SES. Households with lower SES cannot, or do not know how to, provide proper nutrition to mothers and children, leading to poor health. Children in poor health (often measured by height or low birthweight) tend to obtain lower levels of education and worse labor market outcomes over their life (7, 8, 33, 34). In a natural experiment, Almond (35) showed strong negative education, health (physical disability), income, and SES effects for cohorts in utero during the 1918 influenza pandemic compared to those conceived before or after the pandemic. In a British cohort study, children born with low weights were found to pass fewer qualifying exams (33). Studies on twins in Norway (36) and the United States (37) showed that low birthweights led to significantly lower height, IQs, education, and earnings. Height was also a strong predictor of obtaining higher education in Sweden (38) and the United States (34). This empirical evidence suggests that part of the positive gradient between education and health originates in the effect of childhood health on educational attainment. The quantitative importance is still
questioned, but it appears that reverse causality can explain, at most, a small proportion of the observed gradient (7).

Other hypotheses regarding reverse causality in the SES-health gradient exist but are less well examined than the child health hypotheses. Individuals in poor health may have lower productivity and higher rates of absenteeism, which results in lower labor force participation, wages, and incomes. Studies in this area tend to measure the consequences of specific diseases on productivity or aggregate economic consequences of productivity loss (39, 40). An alternate explanation is that those with lower life expectancies (poor health) may have higher discount rates. They thus invest less in their future and subsequently have lower education, human capital, and income, as well as higher levels of risky behavior (41-43). Finally, it is possible that high health care costs due to poor health lead to lower disposable incomes, particularly in the absence of health insurance (44).

Unobserved factors → SES and health (unobserved heterogeneity)

As demonstrated in the movement to twin studies to examine the education-health link, failure to identify a causal link between SES and health behaviors and/or health led researchers to the explanation that unobserved characteristics within socioeconomic strata drive health differences (9, 13, 15, 30, 31). The unobserved differences are thought to include genetics, family background, access to health care, differential productivity in the use of health information or health care, differential vulnerability, differential exposure to environmental factors (risky work, unsafe neighborhoods, chemical exposure, air and water quality, etc.) (7, 16), and cumulative effects (e.g., individuals, particularly children, from lower socioeconomic situations experience more health shocks than those from higher socioeconomic situations) (16-20, 33, 34). Again, no conclusive evidence has been produced, and where evidence for a causal link is found, the impact seems small.

THE SES GRADIENT IN HEALTH AND THE AMERICAS AND NONCOMMUNICABLE DISEASES

Few economic studies were identified that focused directly on the SES gradient in NCDs in general or in the Americas. Meghir et al. (15) did examine cancer and CVD mortality rates, but the study population was non-Hispanic whites in the United States. A recent overview of systematic reviews (45) found evidence that supports an association between socioeconomic inequalities and NCDs and risk factors for NCDs, but the overview noted that the evidence is incomplete and is limited by poor methodological quality.

Most well-known studies in the literature use data from the United States or countries of Europe, due to the availability of good-quality data, research capacity, and good publication outlets. The few studies identified that focus on countries in Latin America or the Caribbean tend to be descriptive in nature and are less well known (46-49). An exception is the evaluation of a Mexican program called Progresa (and later Oportunidades), which provided cash transfers to families if their children attended school or medical appointments to receive preventive care (e.g., vaccinations) (50). The assessment found that doubling cash transfers was related to a decrease in stunting, a decrease in body mass index, a lower prevalence of being overweight, and an increase in height for age (50). The experiment was copied in some areas of the United States, but the results were not duplicated (2, 50).

Health data are available in the Americas. Almeida and Sarti (46) discuss cross-sectional datasets collected in Brazil, Chile, Colombia, Jamaica, Mexico, and Peru that could be used to move the study of health and health care inequities forward in the Americas. An analysis of the Costa Rican Longevity and Healthy Aging Study surprisingly found an SES gradient in self-reported health status and healthy years of life but an inverse gradient in CVD and mortality (51). This led the authors to conclude that modern negative health behaviors among high-SES groups in Costa Rica may be reversing CVD risks and thus mortality risk by SES groups.

DISCUSSION

The SES gradient in health is, arguably, one of the most-studied and well-accepted phenomena in health. It is apparent in objective and subjective measures, across virtually all countries, and it is evident at individual and population levels. There is no longer much debate over the relationship between SES and health. However, the literature presented in this piece indicates that exact causal pathways remain elusive. The most promising evidence suggests that child health influences education and thus outcomes in later life (reverse causality), although the magnitude of the effect is still questioned. As well, family background, including income, influences child health, with effects accumulating over the lifespan (income → health). However, despite more than two decades of research, investigators do not fully understand what causes the observed disparities in health across SES. The lack of consistent results across studies could be due to differences in data (including sample sizes), methodologies, institutional settings, unobservables, or a combination of these factors. What is evident is that there is likely no one factor that drives the gradient. Income, education, health care, health behaviors, and other factors have impacts on the gradient. However, the measured impacts of individual factors are often not as large as anticipated.

Economists in public health intervention research are attempting to design research that can examine how multiple intertwined factors (e.g., education, income, the experience of poverty, available health
care resources, etc.) produce individual and interacted health impacts. However, at this point, in light of the mixed results in the literature, policymakers are left with the tough decision of which, if any, of the factors might produce desired results in their populations, given the available resources.

Researchers often call for governments and policymakers to assist in the provision of better longitudinal data to further the research, either through funding or by allowing access to administrative data. Longitudinal data would, for example, help to identify better measures of permanent income, of long-term versus transitory shocks to income and health, and of short-term versus long-term experiences of poverty. It is now possible and becoming less expensive to collect better data on what historically were unobservables (e.g., genetic or family characteristics). These types of improvements in data may assist in the identification of the elusive causal pathways of the health gradient.

Making administrative data (such as registries of mortality, cancer, CVD, and other diseases) available and linking these databases to other administrative databases (such as tax and education files) might provide much improved data to study the links between socioeconomic status and NCDs. Health and sociodemographic data are being collected in some countries of Latin America and the Caribbean, and some interesting experiments have been undertaken (e.g., Progrésa/Oportunidades).

Building interest in funding more experiments and/or collecting better data would help increase the research capacity in the Americas and elsewhere. In turn, better research capacity could provide several benefits. It would enable researchers to apply the economic theories and methodologies that have been used in many of the studies cited in this piece. It would help to identify new methods to detect the causal pathways needed to address SES inequalities in health. And, it would provide policymakers with the information needed to make evidenced-based policy decisions in the Americas and elsewhere.

**Conflicts of interest.** None declared.

**Disclaimer.** Authors hold sole responsibility for the views expressed in the manuscript, which may not necessarily reflect the opinion or policy of the RPSP/PAJPH or PAHO.

**REFERENCES**

Opinion and analysis

Curtis • Causal explanations for socioeconomic inequalities in health

50. Fawley BW, Juvenal L. Mexico’s Oportunidades program fails to make the Grade in NYC. Reg Econ. 2010;18(3):10–1.

Manuscript received on 15 October 2017. Revised version accepted for publication on 22 January 2018.
**RESUMEN**

Una perspectiva económica sobre las explicaciones causales de las desigualdades socioeconómicas en materia de salud

La desigualdad socioeconómica, o el gradiente según la situación socioeconómica, es posiblemente uno de los fenómenos más estudiados en el campo de la salud. El gradiente de la salud es evidente en mediciones objetivas y subjetivas, en prácticamente todos los países y tanto a nivel individual como poblacional. Ya no se debate mucho la relación entre la situación socioeconómica y la salud. Sin embargo, las vías causales exactas siguen siendo difíciles de definir. A fin de promover políticas enérgicas que reduzcan o eliminen el gradiente socioeconómico de la salud, es necesario entender las vías causales, de la intervención al resultado. Si bien los economistas no están convencidos de que se conozcan suficientemente las vías causales del gradiente socioeconómico de la salud, han producido un volumen sustancial de trabajo a partir del cual avanzar. En este artículo se comentan brevemente los fundamentos teóricos usados por los economistas como base para estudiar las vías causales del gradiente de salud. Luego se brinda un panorama conciso de algunos de los datos científicos generados por los economistas. El artículo concluye con una discusión de cómo pueden usarse los datos científicos económicos actuales para ayudar a los responsables de formular políticas a proponer intervenciones que limiten el gradiente socioeconómico en materia de enfermedades no transmisibles.

**Palabras clave**

Desigualdades en la salud; economía.

---

**RESUMO**

Uma perspectiva econômica da explicação causal para as desigualdades socioeconômicas em saúde

A desigualdade socioeconômica, ou o gradiente socioeconômico, é possivelmente um dos fenômenos mais estudados em saúde. O gradiente em saúde é evidente nas medidas objetivas e subjetivas em praticamente todos os países e é evidente ao nível do indivíduo e de população. Já não existe muito debate sobre a relação entre nível socioeconômico e saúde, mas as exatas vias causais continuam mal definidas. Defender uma firme política para reduzir ou eliminar o gradiente socioeconômico em saúde requer conhecer as vias causais, da intervenção ao resultado. Por não estarem convencidos de que existe um entendimento claro razoável das vias causais do gradiente socioeconômico em saúde, os economistas produziram um volume substancial de estudos que servem de base. O artigo aborda resumidamente os princípios teóricos para embasar o estudo das vias causais do gradiente em saúde e apresenta de forma concisa o panorama das evidências geradas pelos economistas. Por fim, se discute como as evidências econômicas atuais podem ser empregadas para ajudar os responsáveis pelas políticas a defender intervenções visando reduzir o gradiente socioeconômico nas doenças não transmissíveis.

**Palavras-chave**

Desigualdades em saúde; economia.