American Journal of Epidemiology **Submitted Manuscript**

Title: Producing change to understand the social determinants of health: the promise of experiments for social epidemiology

Authors: Lisa F. Berkman, Mauricio Avendano, and Emilie Courtin

Correspondence Address: Lisa F. Berkman, Harvard Center for Population and Development Studies and TH Chan School of Public Health, Cambridge MA. (e-mail: lberkman@hsph.harvard.edu)

Affiliations: Harvard Center for Population and Development Studies and TH Chan School of Public Health, Cambridge MA. (Lisa F. Berkman); Center for Primary Care and Public Health (Unisanté), Department of Epidemiology and Health Systems, University of Lausanne, Lausanne, Switzerland (Mauricio Avendano); Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London UK. (Emilie Courtin)

Funding: Medical Research Council MR/T032499/1

Data Availability Statement: N/A

Thanks: We thank anonymous reviewers for their helpful comments

Conference presentation: N/A

Preprint Information: N/A

Disclaimer: N/A

Conflict of Interest: None

Running Head: Experiments for Social Epidemiology

Key words: social experiments, social epidemiology, social policy

Abbreviations: Randomized Controlled Trials: RCT/ Myocardial Infarction: MI/ Intention-to-treat: ITT

Abstract

In this commentary invited for the 100th anniversary of AJE, we discuss the addition of randomized experiments, along with natural experiments that emulate randomized trials using observational data, as designs in the social epidemiologist's toolbox. These approaches transform the way we define and ask questions about social exposures. They compel us to ask questions about how well-defined interventions change a social exposure that might lead to changes in health. As such, experiments are of unique public health and policy significance. We argue that they are a powerful approach to advance our understanding of how well-defined changes in social exposures impact health, and how credible social policy reforms may be instrumental to address health inequalities. We focus on two research designs. The first is a 'pure' randomized controlled trial (RCT) in which the investigator defines and randomly assigns the intervention. The second one is a natural experiment, which exploits the fact that policies or interventions in the real world often involve an element of random assignment, emulating an RCT. To give the reader our bottom line, while acknowledging their limits, we continue to be very excited about the promise of RCTs and natural experiments to advance social epidemiology.

Randomized experiments, along with natural experiments that emulate randomized trials using observational data, are increasingly common designs in social epidemiology, which have transformed the way in which we define and ask questions about social exposures. They compel us to ask questions about how measurable and actionable changes in social exposures (*e.g.*, an additional year of schooling or a $500 monthly cash transfer) impact health. Observational studies in social epidemiology traditionally focus on identifying social "risks factors", but they rarely aim to identify the causal impact of concrete social policies or interventions. Thus, experiments are of unique public health and policy significance: they offer a powerful approach to advance our understanding of how well-defined social and structural changes impact health, and how social policy reforms may be instrumental to address health inequalities. Our main arguments apply to experiments and quasi-experiments beyond social policies (e.g. natural disasters, political events); however, we focus on social policies because of their relevance for a social epidemiology that can inform action to address health disparities.

Prior literature has articulated excellent conceptual and methodological insights on RCTs and natural experiments in social epidemiology. For example, Diez Roux [1] argues for a practical approach that conceptualises social causes in a continuum from 'upstream' distal factors to specific interventions on narrowly defined factors, where only the latter maybe studied through experiments. Glymour and colleagues offer excellent theoretical and technical resources for understanding key assumptions and technical issues that arise in natural experiments and instrumental variable analysis in particular [2, 3]. In this commentary, we aim to do something different by reflecting on what we have learned in the practice of doing RCTs and using natural experiments to address social epidemiology questions. In particular, we aim to draw attention to

challenges that arise in the design and interpretation of an experiment to test the impact of a social determinant on health; and in the exciting, yet gruelling task of evaluating the impact of policies using natural experiments. As we shall argue, drawing conclusions from experiments on the causal impact of social determinants on health is not straightforward, and requires us to think very carefully about the specific nature of the intervention tested in an RCT, or specificities of social policies evaluated in natural experiments. More often than not, we have identified unintended consequences of interventions tested with an experimental design, which result from not thinking through some pitfalls in design or analysis. This commentary focuses on our experiences in the field and some lessons learned from counterintuitive results.

In this essay, we focus on two research designs. The first is a 'pure' randomized controlled trial (RCT) in which the investigator defines and randomly assigns the intervention. The second one is a natural experiment, which exploits the fact that policies or interventions in the real world often involve an arbitrary cut-off assigning people to 'treated' and 'control' groups, thereby emulating an RCT. To give the reader our bottom line, we continue to be very excited about the promise of RCTs and natural experiments to advance social epidemiology. At the same time, we acknowledge the limits of experimental work to address certain types of questions that are best suited to observational studies. Important issues we will cover in this piece relate to: (1) the length of exposure needed to measure health impacts as well as the identification of appropriate etiologic periods; and (2) the tension between evaluating the impact of a narrow or specific intervention *vs.* identifying impacts of long-run social exposures on health. The effect of exposures such as structural racism or poverty accumulate over long-periods and affect multiple dimensions of wellbeing, e.g., employment, health. Experiments are better suited to assess

'acute' interventions (*e.g.*, income transfers) or specific policy changes (*e.g.*, abolition of Jim Crow laws). It is the scope of what they assess that differentiates them, and make the approaches complementary.

Randomized Controlled trials in social epidemiology

After 10-15 years as a social epidemiologist working with observational data from community-based cohort and national studies, I (LB) decided to make an important turn toward experimental approaches in the mid-nineties as a way to understand the impact of social determinants on population health. I did this for two reasons. First, the experimental design rests on a powerful framework for identifying causes of poor health and secondly, but perhaps more honestly and importantly for me, it shows us how to intervene to improve population health. Public health rests on both acts of identifying causal exposures and taking action to address those causes to improve population health and reduce health inequities [4]. Experiments enable us to learn how generating change in social exposures can improve health. They also compel us to identify a theoretically plausible sensitive or critical period in the life-course and test the hypothesis that intervening on that social exposure in that period would lead to changes in health. In many ways, this course correction has been incredibly rewarding; while in other ways, it turned out to be naïve. The shortcomings of RCTs recently highlighted by economists, sociologists and epidemiologists are important to note [5-7]. We focus here on challenges specifically in the context of social epidemiology.

My first foray in experiments was ENRICHD, a multi-center randomized, double blinded clinical trial to reduce depression and improve social support to reduce myocardial infarctions (MI) and mortality in men and women who already had their first MI [8]. It was designed as a psychosocial intervention. The study design was strong, methods impeccable, study population diverse in terms of race, ethnicity and gender. The intervention was based on individual cognitive behavioral therapy oriented to reducing social isolation and/or depression. Each participant received weekly then monthly visits by a professional therapist over a 6 month period. What did we learn from this major psychosocial experiment? On the positive side, we learned we could recruit and maintain a large diverse sample of men and women (*N*=2,481) and that in fact we could modify levels of depression and social support within 3-6 months. However, intention-to-treat (ITT) estimates revealed a null impact of the intervention on major outcomes of interest, although post hoc analyses showed some effects for white men . The fact that we observed a null effect forced us to spend time thinking about why a very comprehensive intervention in a well- designed trial did not lead to expected outcomes.

In this process, we learned several lessons. First, we were very unclear about the etiologic period and the point in the life-course at which the intervention could change levels of depressive symptoms or social support, which in turn would lead to changes in cardiovascular disease. Changing these conditions after someone had their first MI may well have been too late. Experiments gain part of their power not only from assessing causal relationships but also from identifying the etiologic period at which modifying risk will –or will not -modify the outcome. This issue was evidenced in the case of reinfarction and death, as a participants's risk for these outcomes has a small window of high-risk post-MI. In our trial, the mortality rates following MI were undergoing a rapid transition, dropping in risk virtually every year. Our intervention took

months to conduct and,was completed well after this period of highest post MI mortality risk. In terms of identifying events, it seems as if we missed the most important period of risk. With hindsight, it may seem obvious. Yet, it is suprising how these issues continue to plague RCT studies on the impact of social interventions.

A second lesson refers to the importance of understanding the context of a specific RCT, including the barriers participants face. In ENRICHD, our intervention did not contextualize experiences of isolation for those in disadvantaged positions or who faced structural barriers, including structural racism. The model of ENRICHD was an individualized and classically biomedical intervention, rather than a social intervention to reframe the social environment. While we acknowledged social isolation and lack of support was a social experience, we did not orient our intervention to the social environment. We learned that the level of intervention (whether at the individual, group or population level) is a critical decision. For example, the fact that we only observed some effects on white men in ENRICHD, but not in black or Hispanic participants [9], may reflect structural barriers the latter groups face and which limit the success of the intervention to improve health outcomes.

Finally, there is of course the very real possibility that the trial was a more rigorous test of causal relations than an observational study, and neither depression or social isolation is causally related to cardiovascular disease. Given that there are now a number of meta-analyses showing that both social isolation and depression are related to cardiovascular disease and mortality [10, 11], this begs the question of why there are differences between the ENRICHD RCT and observational studies. The substantive limitations outlined above likely play a role, as do analytical differences between RCTs and observational studies [12]. However, this also reiterates our previous point: the impact of a short term and very specific intervention well-suited

to evalution using an RCT design may not offer a full answer to the question of how cumulative exposures that result from long run social isolation and depression impact health.

My second large foray into social experiments was the Work, Family and Health Network (WFHN). Lessons learned from ENRICHD about the importance of the level of intervention suggested that we should focus on organizational level change rather than individual change. The WFHN trial aimed to change the social environment rather than the individual. It did this by intervening with multiple team members through a supervisor level intervention and simultaneously an employee level organizational change team level that gave workers a voice in work place redesign [13]. The organization was identified as the right intervention level through other worksite interventions launched at the time of the WFHN trial [14, 15] as well as pilot data showing that social circumstances could be modified at that level [16]. Results suggest that the intervention improved sleep, reduced tobacco consumption and impacted short-term cardiovascular risks [17-19]. Intervening on the social environment got us closer to the type of social intervention that is likely to lead to a meaningful health change. In subsequent analyses, we also learned that the effect of an intervention can differ dramatically across different industries that operate in vastly different contexts. In particular, the impact of the intervention was often stronger for the IT industry, which was more flexible and better able to implement organizational change for worklife balance and schedule control. The long term care industry, on the other hand, was highly regulated and had less organizational flexibility, affecting our ability to produce change.

In addition to trials we have designed, we have learned a great deal from existing social experiments in the areas of education, income, employment, social assistance or housing. The US has a rich history of testing new social policies using an experimental design [20]. This is an untapped resource for social epidemiologists and recent research has built on these experiments to incorporate health outcomes to their evaluation or to link them to routinely collected health and mortality data [21, 22]. One example is the Opportunity NYC-Family Rewards RCT (2007-2010), which offered cash transfers to low-income families in New York City's most deprived communities conditional on their participation in education, preventative healthcare and parental employment. The intervention led to a large increase in the use of preventative dental care, modest increases in health insurance coverage and parental self-reported health and levels of 'hope' among parents. However, it had no or modest effects on the health of parents and children [23].

A first lesson from Opportunity NYC-Family Rewards is that RCTs do not operate in a vacuum: the NYC intervention operated alongside a range of long-standing social programs, including the Earned Income Tax Credit, Medicaid and substantial housing support. The counterfactual, therefore, was not the absence of financial support, but a rather comprehensive set of social safety net programs, which were minimally enhanced by the addition of cash transfers in the treatment group. This may explain why we observed relatively weaker effects of this conditional cash transfer program in NYC relative to Latin American countires, where these transfers showed stronger benefits, but where the counterfactual was no or limited social benefits. Second, a key feature of Opportunity NYC-Family Rewards was a parental employment incentive. As the Great Recession happened in the middle of the intervention, many adult participants were not able to gain or maintain full-time employment, a key outcome on which

long-term health and well-being effects had been predicted. A third lesson is that experiments draw their conclusion from a narrow definition of change caused by a very specific social intervention. We estimated the effect of a conditional cash transfer received for a period of three years, which is not the same as an intervention that would secure stable income from benefits or employment for a longer period.

To summarize, what we learned from these trials is that they enable us to address critical questions in social epidemiology by pointing to the impact of specific social policy changes that affect the social determinants of health. They also highlight the importance of moving from intervening at the individual level to intervening at the social environmental level at which we theoretically think change can occur. This movement more deeply acknowledges that to produce structural change we need to intervene at this 'structural' level. This is a critical lesson that is now more commonly understood and implemented. Second, we learned that timing is critical both in terms of the etiological period and potentially the length of the intervention. Decisions on timing are made in the absence of complete information but are increasingly informed by pilot studies and early studies in the area. Third, heterogeneity in treatment effects point to the importance of the larger social context in which interventions are implemented, which either enhance or limit the opportunities for success of the intervention. Here we suspect we will always find important and informative variation in effects that will guide future interventions. Fourth, we highlight the importance of understanding the counterfactual –represented by the control group –in interpreting the findings of RCTs. These lessons all point to careful planning and theory-driven thinking that can enhance the success of trials.

Natural experiments to address social epidemiology and policy questions

Natural experiments share with RCTs the fact that the mechanism of treatment assignment is random or quasi-random, but unlike RCTs, researchers have no control on the intervention and they rely on ingenious ways of analysing observational data. The ideal natural experiment is one in which governments or other actors randomly –or quasi-randomly- assign individuals or groups to receive a program or policy. Examples include lotteries used by the US Government for military conscription for the Vietnam war [24] or the state lottery used in Oregon in 2008 to assign low-income residents to get Medicaid insurance [25]. Natural experiments with full reandomisation are rare, however, as policy makers most often target policies based on need. This poses a critical challenge to scientists interested in the impact of policies: as the mechanism of assignment is based on need, treatment and control differ in critical social and economic outcomes to start with. To assess policy impacts, therefore, we cannot simply compare health outcomes of beneficiaries and non-beneficiaries, as these two groups are, by design, not exchangeable.

Natural experiments offer a solution by encouraging scientists to look closely at the policy and identify groups that differ only by virtue of their eligibility to the policy. They are exciting because they enable us to evaluate the impact of policies that have already been scaled-up or implemented, but for which we are not fully aware of their health impacts. They can provide 'real-world' evidence of effectiveness in ways RCTs cannot. On the other hand, they are challenging because they require a detailed understanding of the policy design and implementation, and they rely on our ability to identify a mechanism of treatment assignment deemed as good as random. Some mechanism of random or quasi-random assignment can

nevertheless often be found even for the most elusive policies, as they often affect some, but not all members of a population, and eligibility is often based on arbitrary factors unrelated to the outcome of interest. For example, a policy reform may only affect cohorts born after a presumably random date. The key assumption we make here is that those who are eligible are not different from those not eligible, apart from being subject to the reform. Let us take the example of compulsory schooling laws to illustrate this point and reflect on key lessons we learned from over a decade of using these approaches to answer social epidemiology questions. Compulsory schooling laws have been widely used in economics and social epidemiology to study causal effect of law-mandated changes in schooling duration on a range of social and health outcomes [26, 27]. They 'tick' all boxes of 'good' natural experiments: well-defined policy measures, with a clear threshold for eligibility based on a variable (date of birth) that is well measured in surveys. We and others have exploited the fact that we do not expect people born just before cut-offs for eligibility to longer schooling to be different from those born just after, apart from exposure to the policy. What have we learned from these studies?

The first lesson is that results from natural experiments can differ from results from observational studies using 'conventional' regression methods. Decades of observational research has shown that more years of schooling and higher levels of education are associated with better mental and physical health in adulthood. However, a concern is that it is not education, but other factors associated with educational attainment (*e.g.* family background, cognitive ability, or genetics) that explain why people with higher education have better health. In our work, we tried to address this issue by analysing long-term effects on health of a compulsory schooling reform implemented in France [28, 29]. Implemented in the 1950's, this

reform increased duration of compulsory schooling by two years for cohorts born after January 1953 (treatment group) but left unchanged the years of compulsory schooling for cohorts born before that date (control group). We found that the reform increased years of schooling as intended but its health effects were mixed. It improved cognitive function among men, but it increased depressive symptoms among women [28] and led to poorer biological profiles in respondents who were born in low-income families [28]. These findings are puzzling: Shall we conclude that more years of schooling can lead to poorer health?

An alternative explanation brings us to the second lesson: as we are using historical policy changes, we do not get to design policy reforms and are left with what was decided at the time, however imperfect or difficult to measure that may be. In the case of the French reform, further research on the policy showed that it increased time spent in class but not much more as the reform was not supplemented with broader infrastructural investments or curriculum reform. It did not increase likelihood of getting a higher degree and did not translate into higher earnings or rates of employment in adulthood, contrary to reforms in other countries. Increasing the *quantity* of education alone may not have been effective, and it may have been counterproductive for some groups [30]. In work using a similar reform in the United Kingdom, we found that young people who were forced to stay at school longer but wished they could have left early had worse mental health across the life course than young people who were also required to stay longer but wanted to stay at school [31]. The reform negatively affected individuals who felt they lost a year of work experience or vocational training. Of course, these policies may have different health effects for historical cohorts than they would if they were implemented today in the same country. For example, in the French reform, women eligible to longer schooling faced

considerable challenges to converting additional years of schooling into better diplomas and jobs in the late 20th Century. However, women benefitting from the same policy in the 21st Century may be better able to reap the benefits of longer schooling, given women's advancement in the labor market during the last decades. Again, context and access to other resources matters. The literature on compulsory schooling reforms illustrate the trade-offs between increasing our ability to make causal claims by studying a well-defined policy intervention; and losing sight of the bigger picture by focusing on a very specific (and limited) way to intervene on education.

A third lesson is the importance of interdisciplinary collaborations with the social sciences, economics, social policy and law. This is also true for the investigation of the health effects of social experiments, as we learned in the NYC Family Rewards program discussed above. Interdisciplinary collaborations are essential to understand the theory, mechanisms and implementation of historical social policies. For example, collaboration with economists was essential in order to understand the complex social and economic impacts of compulsory schooling laws [28, 29, 31], the earned income tax credit [32] and maternity leave benefits [33], and their consequence for health. Interdisciplinarity is critical because natural experiments require a full understanding of social policy reform in order to identify a control group that offers a good counterfactual. Identifying a good control group requires expert knowledge of the laws underlying a policy change and historical research on how the policy was implemented. Another dimension of these necessary collaborations is that the implementation of quasi-experimental methods requires a good understanding of causal inference theory and econometric approaches. Training is a key challenge that we have encountered in our own journey trying to analyze these policy reforms and their consequences for health, especially as the field evolves rapidly as

econometric methods become more and more sophisticated. Our own training was often *ad hoc*, and we were lucky to work in interdisciplinary departments. We hope that approaches such as regression discontinuity designs, instrumental variables and difference-in-differences (and importantly their assumptions and limitations) will become a mainstay of social epidemiology training.

Finally, although natural experiments rely on the strength of the research design, a fourth lesson we have learned is that a strong design is not enough, logistics are also key. For example, a natural experiment requires that we have data at the right time and for the right populations. Often, this means data before and after a policy was enacted and implemented; or information that enables us to distinguish individuals affected by the reform from seemingly comparable individuals who were not affected by the reform.

Bearing in mind these challenges, we cannot but emphasise the huge value of natural experiments. They offer a rigorous, yet pragmatic way to understand how policies impact the health of populations. Under certain assumptions, they enable us to study the causal impact of policies in 'real life', unlike RCTs, which often requires us to deliver policies in limited ways that differ from real world practices. They also enable us to evaluate the impact of policies that we think may have negative effects on health and inequalities but cannot be randomised for ethical, political or practical reasons, such as mass incarceration or welfare benefit cuts. Importantly, even if only partially, they also enable us to study the impact of deep societal changes, including those arising from major social transformations such as the French revolution [34], the fall of the Berlin wall [35] or exposure to armed conflict [36].

Conclusions: the strengths and limitations of experiments for social epidemiology

We reflect on the value of experiments in social epidemiology as the AJE celebrates its 100th anniversary. Looking back at our own work and the work of others, it is clear to us that social experiments are a powerful research design to test novel hypotheses on the impact of changing social determinants of health through interventions and policies. They enable us to evaluate health impacts of many social and economic policies that were not explicitely designed to improve health. Their spillover effects on health have often been unrecognized and therefore not "counted" as real benefits (or costs). Yet, experiments need to be grounded in social theory and build on observational research that point to theoretically plausible social exposures, interventions, mechanisms, and etiologic periods. The interplay between theory, observational research and experiments is therefore critical to advance social epidemiology.

There is a danger in going too far in making causal claims based on experiments that test interventions that have weak or modest effects on the very social determinants whose health effects we are interested to understand, such as poverty or education. In these situations, we are unlikely to get a full picture of the complex causal mechanisms by which these social determinants affect health. Using experiments to advance knowledge on how social and economic factors influence health would require us to have more powerful interventions that can effectively change social and economic determinants. This is not a limitation of experiments per se. Instead, it is a limitation of our knowledge base on effective social policies and interventions that generate the transformative changes we need to produce to understand the 'full' impact of

social and economic exposures on health. Often, we overreach here as we extend "meaning" from a very specific intervention to a larger social condition.

This leads us to a critical distinction between experimental and observational research in social epidemiology: examining the health impact of a change in a social determinant (e.g., income, education) produced by a social policy or intervention is not the same as examining the impact of that social determinant on health. Social policies and interventions produce, in most cases, only a modest change in social determinants. This begs the following question: In which case are experiments useful to study the health impact of social determinants? conceptualising social causes as a continuum, Diez Roux [1] argues that experiments are appropriate to study the impact of a narrowly defined "proximal" factor on which we can intervene, but less useful to assess the impact of 'upstream' distal factors. Our argument is somewhat different: we argue that social causes might be better distinguished by the breadth of social experience and time of exposure rather than proximal *vs*. distal. Back to our example on compulsory schooling laws, we are approaching the effect of education through a one or two year increase in the duration of schooling in adolescence, which cannot reveal the impact of a more transformative change that would profoundly modify complex educational trajectories and experiences. This is even more critical when we try to study social exposures like structural racism using these tools. Changing Jim Crow laws, while focusing on a structural and 'distal' determinant (the law), is a narrower and more specific exposure than decade-long exposure to structural racism. The risk is that we focus on what we can measure rather than what we should measure, akin to looking for the keys only under the lamppost because that is where the light is. There is, in our view, a need for

bolder experiments that dare to test novel and transformative interventions that may one day help us to identify ways to meaningfully reduce health inequalities.

In conclusion, social epidemiologists would do well to embrace the exciting opportunities that that experiments have to offer to understand how interventions that generate changes in social exposures lead to changes in health. Social experiments might best be understood as one piece of the arc we are trying to build with robust evidence of the social determinants of health. We should also be aware of the challenges of experiments. For RCTs, critical challenges include identifying the right level, timing and eticological period of intervention; understanding the larger social context; and being clear about the counterfactual. For natural experiments, we highlight the challenge of reconciling results from observational studies and natural experiments, and the limitations of looking at historical policy changes in particular contexts. For all experiments, we also emphasize the critical importance of interdisciplinary collaboration. Overall, experiments offer a critical tool, which together with existing theoretical and empirical knowledge, help us to advance our understanding of the complex causal mechanisms linking social determinants and health.

References

1. Diez Roux, A.V., *The Unique Space of Epidemiology: Drawing on the Past to Project Into the Future.* Am J Epidemiol, 2019. **188**(5): p. 886-889.
2. Glymour, M.M., S. Walter, and E. Tchetgen Tchetgen, *Natural experiments and instrumental variables analyses in social epidemiology*, in *Methods in Social Epidemiology*, J.M. Oakes and J.S. Kaufman, eds. San Francisco, CA: Jossey-Bass; 2017:493-538.
3. Glymour, M.M., *Policies as tools for research and translation in social epidemiology*, in *Social Epidemiology*, L.F. Berkman, I. Kawachi, and M.M. Glymour, eds. New York, NY: Oxford University Press; 2014:452-478.
4. Galea, S. and M.A. Hernán, *Win-Win: Reconciling Social Epidemiology and Causal Inference.* American Journal of Epidemiology, 2020. **189**(3): p. 167-170.
5. Deaton, A. and N. Cartwright, *Understanding and misunderstanding randomized controlled trials.* Soc Sci Med, 2018. **210**: p. 2-21.
6. Schwartz, S., et al., *Is the "well-defined intervention assumption" politically conservative?* Social science & medicine (1982), 2016. **166**: p. 254-257.
7. Krieger, N. and G. Davey Smith, *The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology.* International Journal of Epidemiology, 2016. **45**(6): p. 1787-1808.
8. Writing Committee for the ENRICHD Investigators, *Effects of Treating Depression and Low Perceived Social Support on Clinical Events After Myocardial InfarctionThe Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) Randomized Trial.* JAMA, 2003. **289**(23): p. 3106-3116.

9. Schneiderman, N., et al., *Psychosocial treatment within sex by ethnicity subgroups in the Enhancing Recovery in Coronary Heart Disease clinical trial.* Psychosom Med, 2004. **66**(4): p. 475-83.

10. Holt-Lunstad, J., et al., *Loneliness and social isolation as risk factors for mortality: a meta-analytic review.* Perspectives on psychological science, 2015. **10**(2): p. 227-237.

11. Rico-Uribe, L.A., et al., *Association of loneliness with all-cause mortality: A meta-analysis.* PLOS ONE, 2018. **13**(1): p. e0190033.

12. Lodi, S., et al., *Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples.* American Journal of Epidemiology, 2019. **188**(8): p. 1569-1577.

13. Kelly, E.L., et al., *Changing Work and Work-Family Conflict: Evidence from the Work, Family, and Health Network.* American Sociological Review, 2014. **79**(3): p. 485-516.

14. Sorensen, G., et al., *A model for worksite cancer prevention: integration of health protection and health promotion in the WellWorks Project.* Am J Health Promot, 1995. **10**(1): p. 55-62.

15. Sorensen, G., et al., *Double jeopardy: workplace hazards and behavioral risks for craftspersons and laborers.* Am J Health Promot, 1996. **10**(5): p. 355-63.

16. Hammer, L.B., et al., *Clarifying work-family intervention processes: the roles of work-family conflict and family-supportive supervisor behaviors.* J Appl Psychol, 2011. **96**(1): p. 134-50.

17. Hurtado, D.A., et al., *Effects on cigarette consumption of a work–family supportive organisational intervention: 6-month results from the work, family and health network study.* Journal of Epidemiology and Community Health, 2016. **70**(12): p. 1155.

18. Berkman, L.F., et al., *Work–family conflict, cardiometabolic risk, and sleep duration in nursing employees.* Journal of Occupational Health Psychology, 2015. **20**(4): p. 420-433.

19. Buxton, O.M., et al., *Work-Family Conflict and Employee Sleep: Evidence from IT Workers in the Work, Family and Health Study.* Sleep, 2016. **39**(10): p. 1871-1882.

20. Courtin, E., et al., *Can Social Policies Improve Health? A Systematic Review and Meta-Analysis of 38 Randomized Trials.* The Milbank Quarterly, 2020. **98**(2): p. 297-371.

21. Muennig, P., G. Johnson, and E.T. Wilde, *The Effect of Small Class Sizes on Mortality Through Age 29 Years: Evidence From a Multicenter Randomized Controlled Trial.* American Journal of Epidemiology, 2011. **173**(12): p. 1468-1474.

22. Muennig, P., Z. Rosen, and E.T. Wilde, *Welfare Programs That Target Workforce Participation May Negatively Affect Mortality.* Health Affairs, 2013. **32**(6): p. 1072-1077.

23. Courtin, E., et al., *Conditional Cash Transfers And Health Of Low-Income Families In The US: Evaluating The Family Rewards Experiment.* Health Aff (Millwood), 2018. **37**(3): p. 438-446.

24. Angrist, J.D., *Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.* The American Economic Review, 1990. **80**(3): p. 313-336.

25. Finkelstein, A., et al., *The Oregon Health Insurance Experiment: Evidence from the First Year*.* The Quarterly Journal of Economics, 2012. **127**(3): p. 1057-1106.

26. Hamad, R., et al., *How and why studies disagree about the effects of education on health: A systematic review and meta-analysis of studies of compulsory schooling laws.* Soc Sci Med, 2018. **212**: p. 168-178.

27. Lleras-Muney, A., *The Relationship Between Education and Adult Mortality in the United States.* The Review of Economic Studies, 2005. **72**(1): p. 189-221.

28. Courtin, E., et al., *Longer schooling but not better off? A quasi-experimental study of the effect of compulsory schooling on biomarkers in France.* Social Science & Medicine, 2019. **220**: p. 379-386.

29. Courtin, E., et al., *Long-term effects of compulsory schooling on physical, mental and cognitive ageing: a natural experiment.* Journal of Epidemiology and Community Health, 2019. **73**(4): p. 370.

30. Glymour, M.M. and J.J. Manly, *Compulsory Schooling Laws as quasi-experiments for the health effects of education: Reconsidering mechanisms to understand inconsistent results.* Social science & medicine (1982), 2018. **214**: p. 67-69.

31. Avendano, M., A. de Coulon, and V. Nafilyan, *Does longer compulsory schooling affect mental health? Evidence from a British reform.* Journal of Public Economics, 2020. **183**: p. 104137.

32. Courtin, E., et al., *Effect of Expanding the Earned Income Tax Credit to Americans without Dependent Children on Psychological Distress (Paycheck Plus): a Randomized Controlled Trial.* Am J Epidemiol, 2021.

33. Avendano, M., et al., *The long-run effect of maternity leave benefits on mental health: Evidence from European countries.* Soc Sci Med, 2015. **132**: p. 45-53.

34. Acemoglu, D., et al., *The Consequences of Radical Reform: The French Revolution.* American Economic Review, 2011. **101**(7): p. 3286-3307.

35. Frijters, P., J.P. Haisken-DeNew, and M.A. Shields, *Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification.* American Economic Review, 2004. **94**(3): p. 730-740.

36. Akbulut-Yuksel, M., *War during childhood: The long run effects of warfare on health.* Journal of Health Economics, 2017. **53**: p. 117-130.