Max-Planck-Institut für demografische Forschung

Max Planck Institute for Demographic Research

Konrad-Zuse-Strasse 1 ▪ D-18057 Rostock ▪ Germany ▪ Tel +49 (0) 3 81 20 81 - 0 ▪ Fax +49 (0) 3 81 20 81 - 202 ▪ www.demogr.mpg.de

MPIDR Working Paper WP 2019-004 I February 2019
(Revised October 2019)

A Generalized Counterfactual Approach to Decomposing Differences Between Populations

Nikkil Sudharsanan
Maarten J. Bijlsma I bijlsma@demogr.mpg.de

This working paper has been approved for release by: Kieron Barclay (barclay@demogr.mpg.de), Deputy Head of the Laboratory of Population Health.

A Generalized Counterfactual Approach to Decomposing Differences Between Populations

Nikkil Sudharsanan[a] and Maarten J. Bijlsma[b]

[a]Heidelberg Institute of Global Health
Heidelberg University
Im Neuenheimer Feld 130.3
69120 Heidelberg
Germany
Tel: +49 176 59587870
Fax: +49 0622 1565948
nikkil-sudharsanan@uni-heidelberg.de

[b]Laboratory of Population Health
Max Planck Institute for Demographic Research
Konrad-Zuse-Straße 1
18057 Rostock
Germany
Tel: +49 381 2081-211
bijlsma@demogr.mpg.de

Abstract

One central aim of the population sciences is to understand why one population has different levels of health and well-being compared to another. Various demographic and regression decompositions have been used to decompose population-differences in a wide range of outcomes. We provide a way of implementing an alternative decomposition method that, under certain assumptions, adds a causal interpretation to the decomposition by building upon counterfactual-driven methods. Our approach has the advantage of flexibility to accommodate different types of outcome variables and any summary population measure. By using Monte Carlo methods, our approach does not rely on closed-form approximate solutions and can be applied to any parametric model without having to derive any decomposition equations. We demonstrate our approach through two motivating examples using data from the 1970 British Birth Cohort Study and the Korean Longitudinal Study of Aging. Our first example decomposes socioeconomic status differences in three different summary measures of fertility and our second addresses the classic demographic question of the contribution of smoking to sex differences in life expectancy. Together, our two examples outline how to implement a very generalized decomposition procedure that is theoretically grounded in counterfactual theory but still easy to apply to a wide range of situations. We provide example R-code and an R-function [package in development].

Keywords: decomposition, causal inference, Monte Carlo, parametric g-formula, population models

Introduction

One central aim of the population sciences is to understand why one population has different levels of health and well-being compared to another. Recent examples of this question include understanding why African Americans have worse health compared to white Americans (Geruso 2012; Kittner et al. 1990), why the United States has lower life expectancy compared to other high-income countries (Ho 2013), why poorer individuals in Finland have higher mortality compared to more affluent individuals (Martikainen et al. 2014), and why the southern American states have higher rates of cardiovascular disease compared to other parts of the country (Steckel and Senney 2015). By identifying the sources of differences across populations, these studies provide an important first step for determining what can be done to reduce disparities.

Different disciplines have developed various methods to answer this wide array of questions. Demographic decompositions, such as the Kitagawa (Kitagawa 1955), Arriaga (Arriaga 1984), step-wise (Andreev, Shkolnikov, and Begun 2002), and related decompositions (Chevan and Sutherland 2009; Gupta 1978; Horiuchi, Wilmoth, and Pletcher 2008) use aggregate data to decompose differences between populations. Regression decompositions, such as the Oaxaca-Blinder (OB) decomposition (Blinder 1973; Oaxaca 1973) and its nonlinear extensions (Powers and Yun 2009; Yun 2004), use individual-level data and are employed frequently in economics and sociology. Both these classes of decompositions are based around mathematical identities, where estimates of the contribution of specific characteristics are derived such that they sum to the total difference in the outcome between groups. For this reason, while the decomposition results have clear mathematical interpretations, they often have ambiguous causal interpretations that are not mapped to specific counterfactual scenarios.

Recent advances in epidemiology and psychology provide a new perspective to decompositions by situating them in causal inference and counterfactual theory (Jackson and

VanderWeele 2018; Nandi, Glymour, and Subramanian 2014). Jackson and VanderWeele (JVW) (2018), develop a general decomposition theory where the importance of specific characteristics to differences between populations is evaluated through hypothetical intervention scenarios with clear counterfactual interpretations. For example, as one scenario, JVW estimate the contribution of educational attainment (measured through test scores) to black-white differences in wages by estimating how much the black-white disparity reduces when test scores are intervened upon and brought to the same level as the white population in comparable demographic strata. The idea of a contributing variable being intervened upon is important because it implies that the decomposition results can be biased if there are unobserved confounders of the relationship between the contributing variable and the outcome. For example, if the decomposition was conducted on the basis of unconditional associations or correlations, the resulting estimate of the contribution of test scores would be biased from a causal perspective because it includes both the true effect of intervening on test scores plus the effect of intervening on other factors that are correlated with both test scores and wages (such as childhood socioeconomic status). Importantly, under some conditions, the JVW decomposition is equivalent to the linear Oaxaca-Blinder decomposition methods when all relevant confounders are adjusted for (see JVW for a formal proof).

There are important limitations to existing decomposition methods that constrain their ability to answer questions that are common in demography and the population sciences. The JVW and extensions of the OB decomposition are only built to decompose mean differences between populations. This means they cannot be applied to many important summary population measures, such as life expectancy, total fertility rates (TFR), and age-standardized prevalence rates. The OB and JVW decompositions also generally require that the researcher separately derive a new set of equations depending on the distribution of the outcome variable of interest, and in many cases, may require making approximations to obtain closed-form solutions. This limits their applicability and use by a wider scientific community. Many demographic decompositions, such as the step-wise decomposition algorithm and line-integral decomposition are formulated to overcome these two limitations by providing general algorithms for decomposing any function, not just mean contrasts (Andreev et al. 2002; Horiuchi et al. 2008). However, because these decompositions are based around mathematical identities instead of causal frameworks, the resulting estimates often do not map to specific counterfactual intervention scenarios. Related to this issue is that the decompositions are mostly non-parametric and can therefore only account for confounding through stratification. This rapidly creates dimensionality problems when there are multiple confounders or confounders with many strata.

In this paper, we implement a micro-data-based counterfactual decomposition that is easily applied to a wide range of questions common in demography and the population sciences. Our approach uses parametric models and Monte Carlo estimation to extend existing decomposition methods in two ways. First, building on the OB and JVW decompositions, our approach is not limited to just mean differences and can be applied to decompose any contrast of any summary population measure, such as life expectancy, prevalence ratios, and even distributional differences such as quantiles (incl. medians). Second, our implementation does not require the analyst to derive decomposition equations and can be estimated flexibly for different types of outcome and explanatory variables. Third, our approach extends classic demographic decompositions by using parametric models to control for confounders and thus avoids the need for stratification and problems created by high dimensionality. The method does not come without tradeoffs however. In contrast to the OB and JVW decompositions, our approach requires substantial computational power and time, and in contrast to demographic decompositions, which are often based on widely accessible aggregate data, our approach requires large scale individual-level data and parametric modeling assumptions.

In the first section of our paper, we introduce the counterfactual theory and background of our approach using the motivating example of decomposing socioeconomic status differences in fertility in the United Kingdom. We next demonstrate the flexibility of our approach by addressing a classic demographic question: "what is the contribution of smoking to sex differences in life expectancy in South Korea?" Collectively, our two examples encompass multiple outcomes and mediating variables common in the population sciences and demonstrate how to implement a very generalized decomposition procedure that is theoretically grounded in counterfactual theory but still easy to apply to a wide range of situations.

A Counterfactual Approach to Decomposition

Concepts

We motivate and develop our approach through the question, "what is the contribution of years spent in schooling to childhood socioeconomic status (SES) differences in fertility in the United Kingdom?" We hypothesize that part of the reason behind why women with higher childhood SES (henceforth "high SES women") have lower fertility compared to women with lower childhood SES (henceforth "low SES women") is that their fertility was delayed because they spent more years in school. To begin answering this question, our approach requires specifying a concrete definition for the "contribution" of years spent in schooling to differences in fertility. We adopt a counterfactual perspective and ask, "how large would the difference in fertility be if both high and low SES women completed the same number of years of school?" Based on this counterfactual, the contribution of schooling is revealed by seeing how much childhood SES differences in fertility change when we intervene to equalize schooling between high and low SES women.

The first step needed to estimate this counterfactual is to specify exactly what level of schooling we are equalizing high and low SES women to. A number of possible options exist. For example, we could set schooling for both groups to secondary schooling levels (12 years), set the low SES groups to have the same distribution of schooling as the high SES group, or vice versa. When the relationship between an outcome (such as fertility) and a mediator (such as schooling) is linear, the choice of this reference distribution does not affect the contribution estimate. However, when the outcome and mediator have a nonlinear relationship, the choice is non-arbitrary and different distributions can result in different contribution estimates (Andreev et al. 2002). For this reason, the choice of the reference distribution should be informed by substantive concerns (e.g. what makes sense from a policy perspective?) and inferential concerns (e.g. certain values may be outside the range observed in the data and should therefore be avoided). For our example, we set the low SES women to have the same distribution of schooling as the high SES women, since this maps to a clear intervention of "how much would SES differences in fertility change if we intervened to improve schooling among low SES women?"

The second main step is to specify a summary population measure. We could use a simple mean as our summary measure and compare the mean number of children between high and low SES women. We are not just limited to the mean, however, and could also examine other summary measures such as median time to first birth, or even a function of means such as the total fertility rate (TFR).

After specifying a summary measure, the third main step is to re-estimate what the summary measure of fertility would be for low SES women after they have been assigned a new schooling distribution. We adopt a potential outcomes framework where every woman in our data is assumed to have several potential fertility outcomes corresponding to each value of schooling they could have theoretically attained. In reality, we only ever observe a potential outcome for the actual level of schooling a particular woman attained. For example, for a woman with 12 years of schooling and 3 children, her potential fertility outcome corresponding to 12 years of schooling would simply be 3 children. However, if we reassigned low SES women to have the schooling distribution of high SES women, in the counterfactual world, this same woman may now have "attained" 20 years of schooling.

The main causal inference problem, which we outline more formally in the next section, is to determine what fertility for this woman would be if instead of the 12 years of schooling she actually attained, she instead attained 20 years. This process needs to be done for all low SES women in the data before forming the counterfactual summary measure of fertility.

Based on both steps 1, 2, and 3, we construct our quantitative estimate as a percentage change and formally define "contribution" as:

$$Contribution = 1 - \frac{\Delta S(Y_{counterfactual})}{\Delta S(Y_{observed})} \tag{1}$$

where Y represents fertility, $S(\,)$ represents our chosen summary measure, Δ refers to the difference in the summary measure of fertility between the high and low SES groups, and observed and counterfactual refer to the situations where the distribution of schooling is as empirically observed or where the low SES group has (counterfactually) received the schooling distribution of the high SES group. Note that $S(\,)$ need not just be the mean; a major advantage of our approach is this generality because it allows for the easy comparison of contrasts for multiple summary measures, such as mean or median differences in fertility, or, if Y was an outcome like mortality, a difference of a function of Y such as life expectancy.

In sum, there are three important conceptual steps that must be taken to perform the counterfactual decomposition for this example: (1) setting the low SES population to have the same distribution of schooling as the high SES population; (2) specifying a summary measure; and (3) estimating what the summary measure of fertility would be in the low SES population under the new, counterfactual, distribution of schooling. In the next section, we provide a more formal treatment of the counterfactual perspective and the three decomposition steps.

[**Box 1**: Causal and g-computation terminology about here]

Formal counterfactual approach

Our approach requires that we estimate what fertility (Y) would have been among low SES women (group B) if they were set to have the same distribution of schooling (M) as the high SES women (group A). We first define the potential outcome for an individual when the mediator M is set to a specific value m as $Y(M = m)$. We denote the distribution of M in group A as f_M^A and that in group B as f_M^B and the potential outcome for an individual when the mediator is set to a value drawn from this distribution as $Y(M \sim f_M^A)$ and $Y(M \sim f_M^B)$, respectively. Equalizing M as described (setting the schooling distribution in the low SES women to that of the high SES women), we are now interested in the value of fertility (Y) for individuals in group B when schooling (M) is redistributed to f_M^A: $Y^B(M \sim f_M^A)$).

Next, we need to formally define our summary measure and population contrast of interest. For this exposition, we will use the mean of completed fertility as our summary measure and the difference in this mean between low and high SES women, $E[Y^A] - E[Y^B]$, as our contrast. Given this summary measure and contrast, we are now interested in the mean difference in fertility between SES groups when schooling (M) among low SES women has been redistributed to f_M^A: $E[Y^A] - E[Y^B(M \sim f_M^A)]$. The second term is the counterfactual potential outcome since it is not directly observable in the data. One way to reveal how to estimate this quantity is by expanding the observed mean fertility among low SES women by conditioning on the different values of schooling found in f_M^B:

$$E[Y^B] = \sum_{m \in f_M^B} E[Y|M = m, B] \cdot P(M = m|B) \tag{2}$$

Within this expression, the distribution of schooling (M) for group B, f_M^B, is captured by the set of probabilities, $P(M = m|B)$, for each value of m found in f_M^B. Therefore, if we wanted to estimate what the expected value of Y^B would be if low SES women had the same distribution of schooling as high SES women ($E[Y^B(M\sim f_M^A)]$), we could replace the probabilities of observing each value of M in group B with the corresponding probability of observing that value in group A ($P(M = m|A)$). Then we would estimate the potential outcome as:

$$E[Y^B(M\sim f_M^A)] = \sum_{m\in f_A()} E[Y|M = m, B] \cdot P(M = m|A) \tag{3}$$

This is simply a direct standardization or re-weighting approach to estimating the counterfactual potential outcome.

Unfortunately, in most observational research, this approach will not lead to a correct estimate of the counterfactual average potential outcome since it assumes that the expected value of the outcome Y when M is set to a specific value m_i among those with $m \neq m_i$ can be estimated as the observed expected value for those with $m = m_i$. This condition, known as exchangeability (Greenland and Robins 1986), is often a strong assumption given that there are likely other systematic ways those with different values of M differ that would affect their value of Y. For example, applied to our fertility example, the social and economic characteristics of a region that a woman is born in may affect both the years of schooling that a woman attains (M) and the eventual fertility of that woman (Y). Therefore, if we estimated what fertility for women with low levels of schooling would be if they received more schooling using the observed fertility of those with higher levels of schooling, we would be confounding the true effect of schooling on fertility with the effect of region of birth on fertility. Therefore, in the presence of confounding variables (C), $E[Y^B(M\sim f_M^A)] \neq \sum_{m\in f_M^A} E[Y|M = m, B] \cdot P(M = m|A)$. However, this equality will hold within strata of C:

$$E[Y^B(M\sim f_M^A)|C = c] = \sum_{m\in f_M^A} E[Y|M = m, C = c, B] \cdot P(M = m|C = c, A) \tag{4}$$

This is because within strata, there is no difference in the value of the confounders between those with different levels of schooling. Therefore, differences in stratum-specific potential outcomes are not confounding the effect of schooling with the effect of different confounder values.

We can now estimate $E[Y^B(M \sim f_M^A)]$ by aggregating these conditional potential outcome estimates across the strata of C and M:

$$E[Y^B(M \sim f_M^A)] = \sum_{C} \sum_{f_M^A} E[Y|M = m, C = c, B] \cdot P(M = m|C = c, A) \cdot P(C = c, B) \qquad (5)$$

Estimating this equation amounts to first stratifying by all values of C. Next, within each of these strata, estimating what fertility for low SES women would be if their schooling was re-distributed to the schooling distribution of high SES women in that same stratum. To do this, we would estimate the expected value of fertility for low SES individuals for each value of schooling found in the schooling distribution of high SES women in that same confounder stratum $f_{M|C=c}^A$. We would then multiply these stratum-specific counterfactual-expected fertility values by the share of the stratum with that specific value of schooling in the high SES population $P(M = m|C = c, A)$ and then sum across strata of M and C. This second step matches the distribution between low and high SES women by equalizing the share of women with each value of m in the low SES population to that share in the high SES population (within confounder strata).

At this point, estimating the decomposition first defined in Eq. 1.requires the following three quantities: mean number of children among high SES women $E[Y^A]$, the mean number of children among low SES women $E[Y^B]$, and the mean number of children among low SES women if they had the education distribution of high SES women $E[Y^B(M \sim f_M^A)]$. Inserting these quantities into Eq. (1) leads to our analytic expression for the contribution of schooling to childhood SES differences in fertility:

$$Contribution = 1 - \frac{E[Y^B(M \sim f_M^A)] - E[Y^A]}{E[Y^B] - E[Y^A]} \quad (6)$$

Parametric Modeling and Monte Carlo-Based Estimation

The analytical solution for the counterfactual decomposition (Eq. 6) is simply direct standardization or re-weighting conducted within confounder strata. However, it is challenging to directly estimate this value from observed data for two important reasons. First, the number of expectations that need to be calculated can increase substantially as the number of unique confounder and mediator values increases. This is because the counterfactual potential outcome in Eq. 6 requires estimating separate expectations for each value of the mediators within each stratum of the joint distribution of the confounders. The second major issue that hinders direct estimation is sparsity. As the number of C variables increases, many of the strata will contain few individuals. This creates the following issues: (i) the empirical distribution of M within strata $(P(M = m|C) \; \forall \; m \in \widehat{M}|C)$ will be based off of very few people and therefore may not reflect the underlying distribution from which the values of M are from $f_{M|C}$; (ii) many strata may not contain individuals with the schooling values needed to form the counterfactual expectations (even if these individuals did exist, the resulting expectation will likely have a high variance due to the small within-strata sample size).

Our solution to these two major issues is to use parametric modeling and Monte Carlo based estimation to first parameterize all the expectations, then simulate entire populations under the observed and counterfactual scenarios using the parametric model estimates, and lastly estimate the unconditional expectations by directly taking averages from the simulated data (Bijlsma and Wilson 2019; Robert and Casella 2013). This approach of parameterizing a high-dimensional direct standardization equation is generally referred to as the parametric g-formula and using Monte Carlo estimation and simulations is a frequently applied computational procedure for parametric g-formula estimation (Bijlsma and Wilson 2019; Hernan and Robins 2019; Imai, Keele, and Tingley 2010; Keil et al. 2014; Wang and Arah 2015).

We start by addressing the issue that the observed empirical distribution of M^A within strata of the confounders (C) may not reflect the true strata-specific distribution of M^A due to sparsity. We do this by assuming a distribution type for M^A and fitting a parametric model for M^A as a function of the confounders (C). For example, our main mediator (M) in the fertility example is a count variable (years of schooling); therefore, we might assume a Poisson distribution and then fit the following regression model:

$$g(E[(M^A|C)]) = \alpha_0^A + \sum C_i \cdot \alpha_{C,i}^A \tag{8}$$

where $g(\,)$ is the log-link function.

Next, we use this model to estimate the conditional distribution of schooling among high SES women ($M^A|C$). In the case of a Poisson distribution, which is only characterized by a mean parameter, we estimate the distributions of $M^A|C$ as:

$$\widehat{M}^A|C \sim Poisson(\lambda = \hat{E}[M^A|C] = g^{-1}\left(\hat{\alpha}_0^A + \sum C_i \cdot \hat{\alpha}_{C,i}^A\right) \tag{9}$$

where the hats represent quantities that have been estimated. We can now set the conditional distribution of M in any strata of C for low SES women ($M^B|C$) to that of the same conditional distribution among high SES women ($M^A|C$) by directly drawing new values of schooling (M) for the low SES group from the model-estimated conditional distribution of schooling for high SES women in the same stratum, $\widehat{M}^A|C \sim Poisson\left(\lambda = \hat{E}[M^A|C] = g^{-1}\left(\hat{\alpha}_0^A + \sum C_i^A \cdot \hat{\alpha}_{C,i}^A\right)\right)$. By drawing values from a parametric distribution, this approach side-steps the need to directly estimate the empirical distribution of M^A in any given stratum $P(M = m|C, A)\ \forall\ m \in \widehat{M}_A$.

This approach can be applied to any parametric distribution. For example, if M is binomially distributed, $g(\,)$ would represent a logit link and we could draw values from a binomial distribution.

Similarly, if M is a continuous variable, we would estimate a linear regression and draw from a normal distribution with a mean based on the model and a standard deviation based on the model residuals.

Next, Eq. 6 requires estimates of $E[Y^B|M=m,C]\ \forall\ c \in C$. Similar to the mediator model above, we reduce the dimensionality of this problem by a fitting parametric model for Y^Bas a function of M and C. In the fertility example, our outcome (number of births) is also count distributed and therefore we again assume a Poisson model:

$$g(E[Y^B|M,C]) = \beta_0^B + M \cdot \beta_M^B + \sum C_i \cdot \beta_{C,i}^B \tag{10}$$

Based on this model, the expected value of $Y^B$for any value of M within any of the confounder strata $E[Y^B|M,C]$ is estimated as:

$$\hat{E}[Y^B|M,C] = g^{-1}\left(\hat{\beta}_0^B + M \cdot \hat{\beta}_M^B + \sum C_i \cdot \hat{\beta}_{C,i}^B\right) \tag{11}$$

Therefore, we can now form a parametrized model-based version of $E[Y^B(M{\sim}f_M^A)]$ as:

$$\hat{E}[Y^B(M{\sim}f_M^A)] = \sum_C \sum_M g^{-1}(\hat{\beta}_0^B + M \cdot \hat{\beta}_M^B + \sum C_i \cdot \hat{\beta}_{C,i}^B) \cdot P^*(\widehat{M} = \widehat{m}|C=c,A)) \cdot P(C=c,B) \tag{12}$$

In this equation, $P^*(\widehat{M} = \widehat{m}|C=c,A)$ are the estimates of the conditional distribution of schooling for high SES women (group A), $f_{M|C}^A$, generated by drawing new values of M from a parameterized distribution, $\widehat{M}^A|C \sim Poisson\left(\lambda = \hat{E}[M^A|C] = g^{-1}\left(\hat{\alpha}_0^A + \sum C_i \cdot \hat{\alpha}_{C,i}^A\right)\right)$ rather than directly using the observed empirical distribution of $M|C$ for high SES women $(P(M=m|C,A)\ \forall\ m \in \widehat{M}_A)$.

Now that we have an estimate of $E[Y^B(M{\sim}f_M^A)]$ we could technically estimate the contribution of schooling as defined in Eq. 6 by plugging in $\hat{E}[Y^B(M{\sim}f_M^A)]$, i.e. $Contribution = 1 - \frac{\hat{E}[Y^B(M{\sim}f_M^A)] - E[Y^A]}{E[Y^B] - E[Y^A]}$ since $E[Y^A]$ and $E[Y^B]$ could be directly estimated from the data using the sample means. This is potentially problematic because $\hat{E}[Y^B(M{\sim}f_M^A)]$ is model-based while the other quantities are not. Therefore, differences between the numerator and denominator of Eq. 6 (the counterfactual and observed contrasts) could be due to both the contribution of M and the modeling process. To mitigate model-induced differences, we also estimate $E[Y^A]$ and $E[Y^B]$ using the same modeling process used to form $\hat{E}[Y^B(M{\sim}f_M^A)]$. We use these model-based estimates of the observed data – often referred to as "natural course" estimates – to form our estimates of the contribution of schooling to SES differences in fertility.

To form the natural-course estimates, we begin by estimating the following four models (two for the mediator, and two for the outcome):

$$g(E[(M|C,A]) = \alpha_0^A + \sum C_i \cdot \alpha_{C,i}^A \tag{13}$$

$$g(E[M|C,B]) = \alpha_0^B + \sum C_i \cdot \alpha_{C,i}^B \tag{14}$$

$$g(E[Y|M,C,A]) = \beta_0^A + M \cdot \beta_M^A + \sum C_i \cdot \beta_{C,i}^A \tag{15}$$

$$g(E[Y|M,C,B]) = \beta_0^B + M \cdot \beta_M^B + \sum C_i \cdot \beta_{C,i}^B \tag{16}$$

Note that we assumed separate models for group A and B, with the superscripts indicating that the coefficients need not be the same in both groups. This flexibility comes with the tradeoff of lower precision compared to more constrained models. For example, we could have instead estimated pooled mediator and outcome models with a dummy variable for group and potential interactions of the group dummy with some or all of the C variables. While such models increase precision, they also introduce additional assumptions by constraining the distributions of Y and M across the two groups.

Next, we draw mediator values as before; however, since we are not interested in setting a counterfactual distribution, we draw values for each group based on their own distribution. For example, we would draw values of M for group A (high SES women) based on $\widehat{M}^A|C \sim Poisson(\lambda = \hat{E}[M^A|C] = g^{-1}(\hat{\alpha}_0^A + \sum C_i \cdot \hat{\alpha}_{C,i}^A))$, and values of M from group B (low SES women) based on $\widehat{M}^B|C \sim Poisson(\lambda = \hat{E}[M^B|C] = g^{-1}(\hat{\alpha}_0^B + \sum C_i \cdot \hat{\alpha}_{C,i}^B))$. We then use the outcome models with the natural-course mediator draws to form natural-course estimates of $E[Y^A]$ and $E[Y^B]$:

$$\hat{E}[Y|A] = \sum_{C}\sum_{M} g^{-1}(\hat{\beta}_0^A + \widehat{M} \cdot \hat{\beta}_M^A + \textstyle\sum C_i \cdot \hat{\beta}_{C,i}^A) \cdot P^*(\widehat{M} = \widehat{m}|C = c, A)) \cdot P(C = c, A) \; \#(17)$$

$$\hat{E}[Y|B] = \sum_{C}\sum_{M} g^{-1}(\hat{\beta}_0^B + \widehat{M} \cdot \hat{\beta}_M^B + \textstyle\sum C_i \cdot \hat{\beta}_{C,i}^B) \cdot P^*(\widehat{M} = \widehat{m}|C = c, B)) \cdot P(C = c, B) \; \#(18)$$

Combining these two estimates with the counterfactual estimate of fertility, $\hat{E}[Y^B(M{\sim}f_M^A)] = \sum_C \sum_M g^{-1}(\hat{\beta}_0 + \widehat{M} \cdot \hat{\beta}_M + \sum C_i \cdot \hat{\beta}_i) \cdot P^*(\widehat{M} = \widehat{m}|C = c, A)) \cdot P(C = c, B)$, we can now estimate the contribution of schooling to SES differences in fertility as:

$$Contribution = 1 - \frac{\hat{E}[Y^B(M{\sim}f_M^A)] - \hat{E}[Y^A]}{\hat{E}[Y^B] - \hat{E}[Y^A]} \qquad (19)$$

Other measures than the mean and the pseudo population perspective

For this example, we focused on the mean as our summary population measure and compared mean fertility levels between groups. Our approach is not just limited to the mean, however, and can easily be extended to other summary population measures, such as the median, mode, variance, or a function of population moments such as life expectancy.

For example, suppose rather than the mean number of children we are interested in the contribution of schooling to the difference in median time to first birth between low and high SES women? To estimate this contribution, we begin just as before and start by asking the question "how large would the difference in the median time to first birth between low and high SES women be if we intervened to bring low SES women to the same levels of schooling as high SES women?" Also as before, we fit models for the mediators and set the conditional distribution of schooling for low SES women to that of high SES women by drawing from the parameterized distributions.

At this point for the mean outcome, we used the parametric outcome model with the updated mediator values to estimate the mean value of fertility in each of the strata of the confounders. Using iterated expectations, we then went from this set of conditional means to the unconditional mean fertility outcome. We were able to do this because the mean can be expressed as a weighted sum of conditional means. This process, however, cannot be done for the median, since the overall median does not equal the weighted sum of a set of conditional medians.

Our approach to solving this issue is to introduce another level of Monte Carlo estimation: rather than using the outcome model to predict conditional expectations, we use it to parameterize the distribution of the outcome and then draw individual values of the outcome (time to first birth) from this parameterized distribution in exactly the same way we did for the mediators. We can then estimate any summary measure directly from the simulated data (for example by taking the median or mode of the simulated time to first birth values).

Using these two levels of Monte Carlo estimation (one set of draws for the mediator and another for the outcome), our approach can be understood as estimating the contribution of a mediator or several mediators by first generating an entire counterfactual micro-population where the mediator (and consequently the outcome) is changed in some way and then comparing summary measures from this population to a natural-course population where the mediators have not been changed. This pseudo-population perspective is powerful because it easily allows for comparisons of any contrast we can think of since we have effectively re-generated entire micro-populations for the observed and counterfactual worlds. Provided that the modeling procedure was flexible enough to allow for subgroup-specific effects, we could also focus on specific subgroups by simply limiting our comparison to specific observations in the pseudo-population.

Although this approach may appear challenging, it can actually be estimated by following a straightforward algorithm:

Step 0: Specify starting decisions

a. Decide on a summary measure.

b. Decide on a contrast.

c. Decide on a reference distribution or group.

Step 1: Estimate relationships in the data

a. Fit regression model(s) for the mediator(s) of interest with the same confounders as the outcome model as covariates.

b. Fit regression model(s) for the outcome with the mediator(s) of interest and confounders of the mediator-outcome relationship as covariates.

Step 2: Form the Natural Course Pseudo-Population.

a. Use the mediator model(s) with observed confounder values to simulate mediator values.

b. Use outcome model(s) together with observed confounder values and simulated mediator values to simulate the outcome. This is the natural-course pseudo-population.

c. Using the natural course pseudo-population, estimate the summary measure for both groups and then form the contrast of interest across groups.

Step 3: Form the Counterfactual Pseudo-Population

a. For the non-reference groups, use the mediator model(s) with observed confounder values to simulate mediator values from the reference group.

b. Use the outcome model(s) together with observed confounder values and simulated mediator values to simulate the outcome. This is the counterfactual pseudo-population.

c. Using the counterfactual pseudo-population, estimate the summary measure for both groups and then form the contrast of interest across groups.

Step 4: Compare the contrast of interest in the natural-course and counterfactual pseudo-populations.

To estimate standard errors and to produce stable estimates of the contribution, we have to address two types of variability. First, since we are drawing values of the mediators and outcomes from probability distributions, the exact values assigned to individuals can change across multiple draws. This results in the estimate of the contribution also changing across draws (known as Monte Carlo error). To reduce this error, we conduct Steps 2 and 3 multiple times, each time drawing a new set of mediator and outcome values. We then construct the contrasts for each draw and then average across all these draws to produce stable natural course and counterfactual estimates, before calculating the contribution in Step 4.

Second, because our results are based on a sample, we need to account for sampling variability. This is especially important for the construction of confidence intervals around the estimates. We use a bootstrap procedure to capture this uncertainty, drawing with replacement a fresh sample of size equal to the original data before step 1, conducting the entire analysis *k* times, and then averaging across the *k* bootstrap samples to obtain a point estimate and using the 2.5% and 97.5% percentiles for the confidence interval.

We provide pseudocode for the fertility example to further clarify these steps and how to implement them with common statistical software (Figure 1). We also provide an R-function for easy implementation of our method [Package in development]

Empirical Example 1: Contribution of years of education to childhood socioeconomic differences in fertility

In this example, we demonstrate the application of our approach to an expanded version of the hypothetical question we used to motivate the paper: what is the contribution of years spent in schooling to childhood SES differences in fertility? We demonstrate how to decompose two different types of outcome variables and three classic demographic summary measures: percent childless, median time to first birth (revealing how to apply the method to non-mean-based summary measures), and cohort total fertility rate (TFR).

Data: The 1970 British Birth Cohort

We use data from the 1970 British Cohort Study (BCS70) (Elliott and Shepherd 2006; UK Data Archive 2016). The BCS70 routinely follows around 17,000 individuals born in Great Britain (except Northern Ireland) in a single week in 1970. Beginning in the 26-year follow up, women were asked about their pregnancy history. For this example analysis, we use data on women from the 2008-2009 follow-up wave when the women were 38 years old. We only include those with non-missing baseline and pregnancy follow-up information for a total sample of 3,634 women.

Outcome (Y)

We study three closely related fertility outcomes: being childless at age 38 (binomial), time to first birth from age 16 among those who had at least one child (Poisson), and the number of children born to a woman at age 38 (Poisson). We construct all three outcomes based on self-reported prior pregnancy histories. For this analysis, we only consider live births.

Mediator (M)

Our mediator of interest is the number of years spent in school, based on theory and evidence that greater time spent in schooling could both increase the opportunity cost of having a child and the age at which women give birth (Balbo, Billari, and Mills 2013; Becker 1981; Cleland and Wilson 1987).

Grouping variable

We classify individuals into three childhood socio-economic (SES) groups. To construct the childhood SES groups, we first conduct a principal components analysis on binary indicators for tertiles of parental income when the cohort members were aged 10, tertiles of the respondent's mother's age at delivery, whether their mother was unmarried at the time of delivery, whether the respondent's mother was college educated, whether the respondent's father was college educated, and indicators for occupational classes for both the father and mother. We then create a continuous SES score using the first principal component and classify individuals into tertiles of the score.

Summary measures and contrast

We consider three summary measures corresponding to the three outcomes: percentage of women who are childless at age 38, median time to first birth, and partial-cohort TFR (the mean number of children at age 38; this is partial because women may still have additional children after age 38). For all three summary measures, our contrast of interest is the difference in the summary measure for the low and medium SES groups relative to high SES women. As our counterfactual scenario, we set the education distribution of the low and medium SES women to be equal to that of the high SES women.

Confounders (C)

As confounders of the relationship between years spent in education and childbearing, we adjust for region of birth (Scotland, Wales, Northern England, Midlands, Southern England, and London), the age of the mother of the respondent at first childbearing (less than 22, between 22 and 30, and 30 or more), and the number of siblings that the respondent herself had while living in her parental household.

Models

Mediator models

We model years spent in schooling using the following Poisson regression model:

$$log(E[M|C,SES]) = \alpha_0 + \sum C_i \cdot \alpha_{C,i} + SES_{low} \cdot \alpha_{low} + SES_{middle} \cdot \alpha_{middle}$$

Where M is the count of number of years in education, SES_{low} and SES_{middle} represent dummy variables for low and medium SES categories (leaving high SES as a reference category), and C represents the confounders described previously.

Outcome models

We fit separate models for each of the three outcomes as a function of the years spent in schooling (M), the confounders (C), and SES group:

$$logit\left(E[Y^{childless}|M,C,SES]\right)$$

$$= \beta_0 + M \cdot \beta_M + \sum C_i \cdot \beta_{C,i} + SES_{low} \cdot \beta_{low} + SES_{middle} \cdot \beta_{middle}$$

$$log\left(E[Y^{tfirstbirth}|M,C,SES]\right)$$

$$= \beta_0 + M \cdot \beta_{C,i} + \sum C_i \cdot \beta_i + SES_{low} \cdot \beta_{low} + SES_{middle} \cdot \beta_{middle}$$

$$log\left(E\left[Y^{nchildren}\middle|M,C,SES\right]\right)$$

$$= \beta_0 + M \cdot \beta_{C,i} + \sum C_i \cdot \beta_i + SES_{low} \cdot \beta_{low} + SES_{middle} \cdot \beta_{middle}$$

Note that although we identically named β's in all three models, they are separately estimated and hence represent different quantities. Code for this analysis is available in Supplemental Material will be available with R package. Pseudo-code for calculating the TFR contrast is shown in Figure 1.

Results

Table 1 presents the observed percentage of women who are childless at age 38, the median age at first birth among those that have had at least one birth, and cohort TFR at age 38, across childhood SES groups. Women with higher childhood SES are more likely to be childless, have an older age at first birth, and (consequently) have a lower cohort TFR.

Table 2 presents the counterfactual-based decomposition results for childlessness (Panel A), age at first birth (Panel B), and cohort TFR (Panel C). There is an 8.5 percentage point difference in the natural course difference in percentage childless at age 38 between women with high and low childhood SES, and a 4.2 percentage point difference between women with high and medium childhood SES. After setting the lower two childhood SES groups to have the same distribution of total years spent in education as the high childhood SES group, these differences reduce substantially (5.8 percentage points and 2.1 percentage points for the low and medium childhood schooling groups respectively). Based on this change, we estimate that 32.2% of the difference in childlessness between high and low childhood SES groups and 50.2% of the difference in childlessness between the high and medium childhood SES groups is due to differences in total years spent in education.

We find similarly high levels contributions of schooling to SES differences in median age at first birth. For example, there is a 3.2-year natural course difference in median age at first birth between the high and low childhood SES groups. After setting the low childhood SES group to have the same years spent in education as the high SES group, this difference reduces to 2.6 years. Therefore, we estimate that differences in total years spent in education between the two groups is responsible for 18.7% of the 3.2-year difference in median age at first birth between those with high and low childhood SES.

Lastly, schooling also had a large contribution to SES differences in the cohort TFR at age 38. The difference in TFR between women with high and low childhood SES reduces from 0.27 to 0.19 when total years of schooling for the low SES group is set to that of the high SES group. Therefore, 30.9% of the difference is attributable to schooling differences. The percent contribution for the medium SES group was even larger; while the TFR difference between women with high and medium SES is only 0.10, this reduces to 0.04 when the schooling distributions are equalized, resulting in a 58.2% contribution.

Empirical Example 2: Smoking's Contribution to Sex Differences in Life Expectancy in South Korea

In this example, we demonstrate the application of our approach to a classic demographic question: what is the contribution of smoking to sex differences in life expectancy? For this example, we examine the case of South Korea, a country with particularly large sex differences in both smoking and life expectancy.

Data: The Korean Longitudinal Study of Aging

We use data from the 2006, 2008, 2010, and 2012 waves of the Korean Longitudinal Study of Aging, a nationally representative survey of South Korean adults ages 45 and above (Jang 2015). We focus on adults ages 50 and above for a total sample of 7,615 individuals comprising 500,321 person-month observations.

Outcome (Y)

Our outcome is whether an individual died over the course of the study period. We use information on family-member-reported date of death and the date of last interview for those who did not die to determine the number of person-months that every individual lived between 2006 and 2012.

Mediator (M)

Our primary mediator is a dichotomous indicator for whether an individual reported ever regularly smoking cigarettes.

Grouping variable

We compare men and women ages 50 and above.

Summary measure and contrast

Our summary measure is period life expectancy at age 50. We construct this measure by estimating age-specific mortality rates from the individual-level data and converting these rates into period life expectancies using standard life table techniques (Preston, Heuveline, and Guillot 2000). Our contrast is the difference in life expectancy at age 50 between men and women. As our counterfactual scenario, we set the smoking levels among men to be equal to those among women.

Confounders

We adjust for the following potential confounders of the smoking-mortality relationship: age, how frequently an individual reported drinking alcohol, marital status, schooling, and whether the individual lived in an urban or rural area.

Models

For this example, we estimate pooled outcome and mediator models with an indicator variable for sex. In contrast to the previous example, we also interact the sex variable with the confounders in the mediator model and with the confounders and mediator in the outcome model to allow the conditional mediator and outcome distributions to flexibly vary across men and women.

Mediator model

We parameterize the probability of ever regularly smoking for men and women using the following logistic regression model:

$$logit(E[M|C,SEX]) = \alpha_0 + \sum C_i \cdot \alpha_{C,i} + SEX \cdot \alpha_{SEX} + \sum SEX \cdot C_i \cdot \alpha_{SEXxC,i}$$

Here, M is a binary variable for whether an individual self-reported ever regularly smoking, C represents the vector of confounders described previously, SEX is a dummy variable for female, and the final term represents interactions between the female dummy and the confounders.

Outcome model

Before estimating the outcome model, we convert the data to the person-month level with observations for every month between the initial interview in 2006 to either the date of death or date of last interview. We then model mortality as a function of smoking, sex, and the confounders by fitting the following logistic regression model on the person-month observations (this type of model is also referred to as a discrete failure-time model):

$$logit(E[Y|M,C,SEX]) = \beta_0 + M \cdot \beta_M + \sum C_i \cdot \beta_{C,i} + SEX \cdot \beta_{SEX} + SEX \cdot M \cdot \beta_{SEXxM} + \sum SEX \cdot C_i \cdot \beta_{SEXxC,i}$$

Note that we have also included an interaction between ever smoking and female. By estimating this model on the person-month level, we are able to simulate the precise age at which individuals die. This is needed to correctly estimate age-specific mortality rates. Pseudocode for this example is shown as Figure 2 and code to estimate the example is provided in the supplementary material.

Results

There is a large, 6.7-year difference in life expectancy at age 50 between men and women in South Korea (Table 3). Sex-differences in smoking potentially explain a part of this large mortality difference: 61.0% of men reported ever regularly smoking cigarettes compared to just 4.3% of women.

Figure 3 graphs the natural course and counterfactual life table death distributions and the corresponding estimates of period life expectancy at age 50. In the natural course, men have substantially excess levels of mortality between ages 50 and 80, with especially pronounced differences between ages 70 and 80. In contrast to men, the death distribution for women is concentrated in the oldest ages, with the largest absolute share of life table deaths occurring in the 90+ age group. After setting men to have the same distribution of smoking as women, the large, 6.7-year difference in life expectancy actually reverses, resulting in a 1.8-year advantage for men relative to women. This can be visualized in the counterfactual death distribution for men: the distribution is substantially more concentrated in the older ages even relative to women. The resulting change in the difference in life expectancy corresponds to a (1 - 6.0/(-1.8) = 1.3) 130% contribution of smoking to sex-differences in adult life expectancy in South Korea.

Discussion

We introduce a generalized yet easily applied procedure for decomposing social or population differences in a wide range of outcomes within a counterfactual and potential outcomes framework. This approach is built on Jackson and VanderWeele's counterfactual decomposition theory (Jackson and VanderWeele 2018) and measures the contribution of mediating variables to group differences in a summary population measure by considering how group differences change when the mediating variable is intervened upon in some way. This approach is similar to causal mediation analysis (VanderWeele 2015) with one important difference: whereas causal mediation analysis seeks to split a causal effect into the contribution of mediating pathways, our approach splits an observed association (difference across groups) into the contribution of group differences in the distribution of potential mediators. We demonstrate this approach on two examples that capture questions common in the population sciences: (i) "what is the contribution of years spent in schooling to childhood socioeconomic status differences in fertility in the United Kingdom?"; and (ii) "what is the contribution of smoking to sex differences in life expectancy in South Korea?" In the following sections, we compare our approach to answering these two questions to existing decomposition methods and conclude with a discussion of the general strengths and weaknesses of our method.

Comparison to other methods

Example 1: Schooling and fertility in the United Kingdom

Our first example decomposes differences in three outcomes (childlessness, time to first birth, and number of children) using two summary measures (the mean for childlessness and number of children, and the median for time to first birth). For this example, the major advantage of our approach over existing methods is the ability to decompose median differences; to our knowledge, this is currently not possible with the JVW decomposition theory, regression decompositions such as the Oaxaca-Blinder and its extensions, nor with demographic decompositions.

Although the other two summary measures in the example (mean number of children and percent childless) can be decomposed using existing methods, our approach may still provide a few important advantages. For example, while the JVW theory applies to both summary measures, JVW do not provide decomposition equations for Poisson distributed outcomes (number of children) nor common binary outcomes (childlessness). For these summary measures, our approach can be seen as a way to practically implement JVW's theory to a wider range of outcome types.

Similarly, the mean number of children and percent childless can also be decomposed with the extensions of the Oaxaca-Blinder decomposition developed by Yun and Powers (Powers and Yun 2009; Yun 2004), which uses approximations and Taylor expansions to directly estimate non-linear decomposition equations. Our approach provides an alternative to the Yun and Powers decomposition and uses simulations to avoid needing to derive decomposition equations and making linear approximations. This could be especially helpful when the underlying regressions contain several interaction terms or involve large differences between groups that are not well-approximated by linear equations.

Demographic decompositions, such as the stepwise replacement algorithm (Andreev et al. 2002), line-integral decomposition (Horiuchi et al. 2008), or even the Kitagawa decomposition (Kitagawa 1955), can also be used to estimate the contribution of schooling to group differences in percent childless and mean number of children. The primary limitation of these demographic decompositions is that confounding has to be addressed through stratification. For our example, this means the decompositions would have to be conducted separately in each stratum of the joint density of region of birth (Scotland, Wales, Northern England, Midlands, Southern England, and London), the age of the mother of the respondent at first childbearing (less than 22, between 22 and 30, and 30 or more), and the number of siblings that the respondent herself had while living in her parental household (ranging from 0 to 11) -- representing more than 100 strata. This is likely not possible since many strata would have few to no observations and could not be used to accurately estimate either the counterfactual conditional expectations nor the underlying conditional distributions of schooling (the mediator). On the other hand, if we ignored these variables, the decomposition results may incorrectly estimate the contribution of schooling to fertility differences due to confounding from these other characteristics (see Appendix 1 for an example).

Example 2: Smoking and life expectancy in South Korea

For our second example, we estimate the contribution of smoking to sex differences in life expectancy at age 50 in South Korea. This example highlights a major advantage of our approach over other micro-data-based decompositions: because our summary measure, life expectancy, is not a simple population moment, it cannot be decomposed using the JVW decomposition nor extensions of the Oaxaca-Blinder decomposition.

By contrast, a myriad of demographic approaches have been developed to answer this type of question, most relying on aggregate cause-specific mortality information. Our approach does not replace these decompositions but rather provides a micro-data driven alternative. This may be helpful in circumstances where cause of death data are not available or when the risk factor of interest affects mortality through multiple causes that cannot be solely attributed to the risk factor. This is an important consideration for estimating the contribution of smoking to life expectancy, since smoking affects several different causes of death. Estimating the role of smoking to differences in life expectancy between men and women using traditional decompositions, such as the Arriaga or step-wise replacement algorithm, would therefore require somehow determining the share of multiple causes of death that are attributable to the difference in the prevalence of smoking between sexes.

Our approach bypasses this issue by directly estimating the relationship between the prevalence of smoking and all-cause mortality from individual-level data and using these estimates as the basis for decomposition.

Limitations and disadvantages compared to existing methods

Despite these advantages, our method comes with important trade-offs compared to existing methods. First, compared to the JVW closed-form decomposition equations, regression decompositions such as the Oaxaca-Blinder and its extensions, and many demographic decompositions, our approach requires substantial computational power and time. This is not a trivial consideration and decompositions with large datasets may take hours to even days to complete even when considerable computational power is available.

There are two additional important limitations to our approach when compared specifically to demographic decompositions. The first is that many demographic decompositions use widely available aggregate data sources. Our approach, however, necessitates large, often population-representative, sources of individual-level data. Therefore, our approach is not intended to replace traditional decompositions, but rather to complement these decompositions with more concrete counterfactual interpretations when micro-level data are available. Second, in contrast to the mostly non-parametric approach used in most demographic decompositions (through binning of continuous variables), our approach requires making parametric modelling assumptions that could introduce error if the models do not accurately reflect the real data generating process. To reduce the chance for this error, we strongly recommend that the distribution of simulated outcomes and mediators for each population is compared to the empirical distributions observed in the data. If there are large discrepancies between the simulated and observed values, the model's specification (i.e. assumed outcome distribution or the functional form of covariates) may need to be recalibrated until the differences between the observed and simulated values are no longer large. Although this does not ensure that the models are correctly specified, it helps avoid gross misspecification.

We discuss other considerations for the application and interpretation of our model in Appendix 2.

Conclusions

Decomposing the sources of differences in health and other outcomes is a key research endeavor in demography and other population sciences. We introduce a flexible implementation of the counterfactual decomposition that builds on and generalizes the rich existing body of work on decomposition methods in the health and social sciences. Our approach is a highly flexible and easily implemented way of estimating decompositions that are grounded in potential outcomes and counterfactual theory and applicable to a wide range of population questions.

References

Andreev, Evgueni M., Vladimir M. Shkolnikov, and Alexander Z. Begun. 2002. "Algorithm for Decomposition of Differences between Aggregate Demographic Measures and Its Application to Life Expectancies, Healthy Life Expectancies, Parity-Progression Ratios and Total Fertility Rates." *Demographic Research* 7:499–522.

Arriaga, Eduardo E. 1984. "Measuring and Explaining the Change in Life Expectancies." *Demography* 21(1):83–96.

Balbo, Nicoletta, Francesco C. Billari, and Melinda Mills. 2013. "Fertility in Advanced Societies: A Review of Research." *European Journal of Population/Revue Européenne de Démographie* 29(1):1–38.

Becker, Gary Stanley. 1981. *A Treatise on the Family*. Harvard university press.

Bijlsma, Maarten J. and Ben Wilson. 2019. "Modelling the Socio-Economic Determinants of Fertility: A Mediation Analysis Using the Parametric g-Formula." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 436–455.

Carnegie, Nicole Bohme, Masataka Harada, and Jennifer L. Hill. 2016. "Assessing Sensitivity to Unmeasured Confounding Using a Simulated Potential Confounder." *Journal of Research on Educational Effectiveness* 9(3):395–420.

Chevan, Albert and Michael Sutherland. 2009. "Revisiting Das Gupta: Refinement and Extension of Standardization and Decomposition." *Demography* 46(3):429–449.

Cleland, John and Christopher Wilson. 1987. "Demand Theories of the Fertility Transition: An Iconoclastic View." *Population Studies* 41(1):5–30.

Elliott, Jane and Peter Shepherd. 2006. "Cohort Profile: 1970 British Birth Cohort (BCS70)." *International Journal of Epidemiology* 35(4):836–843.

Geruso, Michael. 2012. "Black-White Disparities in Life Expectancy: How Much Can the Standard SES Variables Explain?" *Demography* 49(2):553–574.

Greenland, Sander and James M. Robins. 1986. "Identifiability, Exchangeability, and Epidemiological Confounding." *International Journal of Epidemiology* 15(3):413–419.

Gupta, Prithwis Das. 1978. "A General Method of Decomposing a Difference between Two Rates into Several Components." *Demography* 15(1):99–112.

Hernan, Miguel A. and James M. Robins. 2019. *Causal Inference*. CRC Boca Raton, FL:

Ho, Jessica Y. 2013. "Mortality under Age 50 Accounts for Much of the Fact That US Life Expectancy Lags That of Other High-Income Countries." *Health Affairs* 32(3):459–467.

Horiuchi, Shiro, John R. Wilmoth, and Scott D. Pletcher. 2008. "A Decomposition Method Based on a Model of Continuous Change." *Demography* 45(4):785–801.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15(4):309–34.

Jackson, John W. and Tyler J. VanderWeele. 2018. "Decomposition Analysis to Identify Intervention Targets for Reducing Disparities." *Epidemiology* 29(6):825–835.

Jang, Soong-Nang. 2015. "Korean Longitudinal Study of Ageing (KLoSA): Overview of Research Design and Contents." *Encyclopedia of Geropsychology* 1–9.

Keil, Alexander P., Jessie K. Edwards, David R. Richardson, Ashley I. Naimi, and Stephen R. Cole. 2014. "The Parametric G-Formula for Time-to-Event Data: Towards Intuition with a Worked Example." *Epidemiology (Cambridge, Mass.)* 25(6):889.

Kitagawa, Evelyn M. 1955. "Components of a Difference between Two Rates." *Journal of the American Statistical Association* 50(272):1168–1194.

Kittner, Steven J., Lon R. White, Katalin G. Losonczy, Philip A. Wolf, and J. Richard Hebel. 1990. "Black-White Differences in Stroke Incidence in a National Sample: The Contribution of Hypertension and Diabetes Mellitus." *Jama* 264(10):1267–1270.

Martikainen, Pekka, Pia Mäkelä, Riina Peltonen, and Mikko Myrskylä. 2014. "Income Differences in Life Expectancy: The Changing Contribution of Harmful Consumption of Alcohol and Smoking." *Epidemiology* 25(2):182–90.

Nandi, Arijit, M. Maria Glymour, and SV Subramanian. 2014. "Association among Socioeconomic Status, Health Behaviors, and All-Cause Mortality in the United States." *Epidemiology* 25(2):170–177.

Neyman, Jerzy S. 1923. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.(Tlanslated and Edited by Dm Dabrowska and Tp Speed, Statistical Science (1990), 5, 465-480)." *Annals of Agricultural Sciences* 10:1–51.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 693–709.

Powers, Daniel A. and Myeong-Su Yun. 2009. "7. Multivariate Decomposition for Hazard Rate Models." *Sociological Methodology* 39(1):233–263.

Robert, Christian and George Casella. 2013. *Monte Carlo Statistical Methods*. Springer Science & Business Media.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688.

Steckel, Richard H. and Garrett Senney. 2015. *Historical Origins of a Major Killer: Cardiovascular Disease in the American South. Working Paper*. 21809. National Bureau of Economic Research.

UK Data Archive. 2016. *1970 British Cohort Study*. University of Essex, Instiute for Social and Economic Research.

VanderWeele, Tyler. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

VanderWeele, Tyler J. and Onyebuchi A. Arah. 2011. "Bias Formulas for Sensitivity Analysis of Unmeasured Confounding for General Outcomes, Treatments, and Confounders." *Epidemiology (Cambridge, Mass.)* 22(1):42–52.

Wang, Aolin and Onyebuchi A. Arah. 2015. "G-Computation Demonstration in Causal Mediation Analysis." *European Journal of Epidemiology* 30(10):1119–1127.

Yun, Myeong-Su. 2004. "Decomposing Differences in the First Moment." *Economics Letters* 82(2):275–280.

Table 1. Observed proportion of women who are childless at age 38, median age at first birth among women who have had at least one birth across childhood SES groups, and Total Fertility Rate, 1970 British Cohort Study 38-year follow-up.

	Low Childhood SES	Middle Childhood SES	High Childhood SES
Childlessness	0.229	0.271	0.314
Age at first birth	25.0	26.0	28.0
Total Fertility Rate	1.72	1.55	1.45

Notes: Childhood SES groups are based on tertiles of a continuous core based on several parental characteristics and parental income when the women were aged 10.

Table 2. Estimates of the contribution of total years of schooling to childhood socioeconomic status differences in every having a birth and median age at first birth among those who have had a birth using the counterfactual decomposition method, 1970 British Cohort Study, 1970-2008.

Panel A: Any birth

	Natural Course Percentage	Counterfactual Percentage	Natural Course Difference	Counterfactual Difference	Percent Mediated
High Childhood SES	0.314	0.314			
	(0.290,0.341)	(0.290,0.341)			
Middle Childhood SES	0.271	0.293	0.042	0.021	50.2%
	(0.246,0.299)	(0.266,0.322)	(0.001,0.078)	(-0.019,0.057)	(14.1%,297.3%)
Low Childhood SES	0.229	0.256	0.085	0.058	32.2%
	(0.204,0.252)	(0.229,0.282)	(0.052,0.120)	(0.021,0.095)	(17.4%,61.7%)

Panel B: Age at first birth among those who had a birth

	Natural Course Median Age	Counterfactual Median Age	Natural Course Difference	Counterfactual Difference	Percent Mediated
High Childhood SES	28.0	28.0			
	(27.9,28.3)	(27.9,28.2)			
Middle Childhood SES	26.0	26.7	-2.0	-1.3	34.8%
	(25.9,26.2)	(26.1,27.0)	(-2.3,-1.8)	(-1.9,-1.0)	(7.4%,50.0%)
Low Childhood SES	24.9	25.4	-3.2	-2.6	18.7%
	(24.2,25.0)	(25.0,26.0)	(-3.9,-3.0)	(-3.1,-2.0)	(5.1%,33.1%)

Panel C: Cohort Total Fertility Rate (TFR)

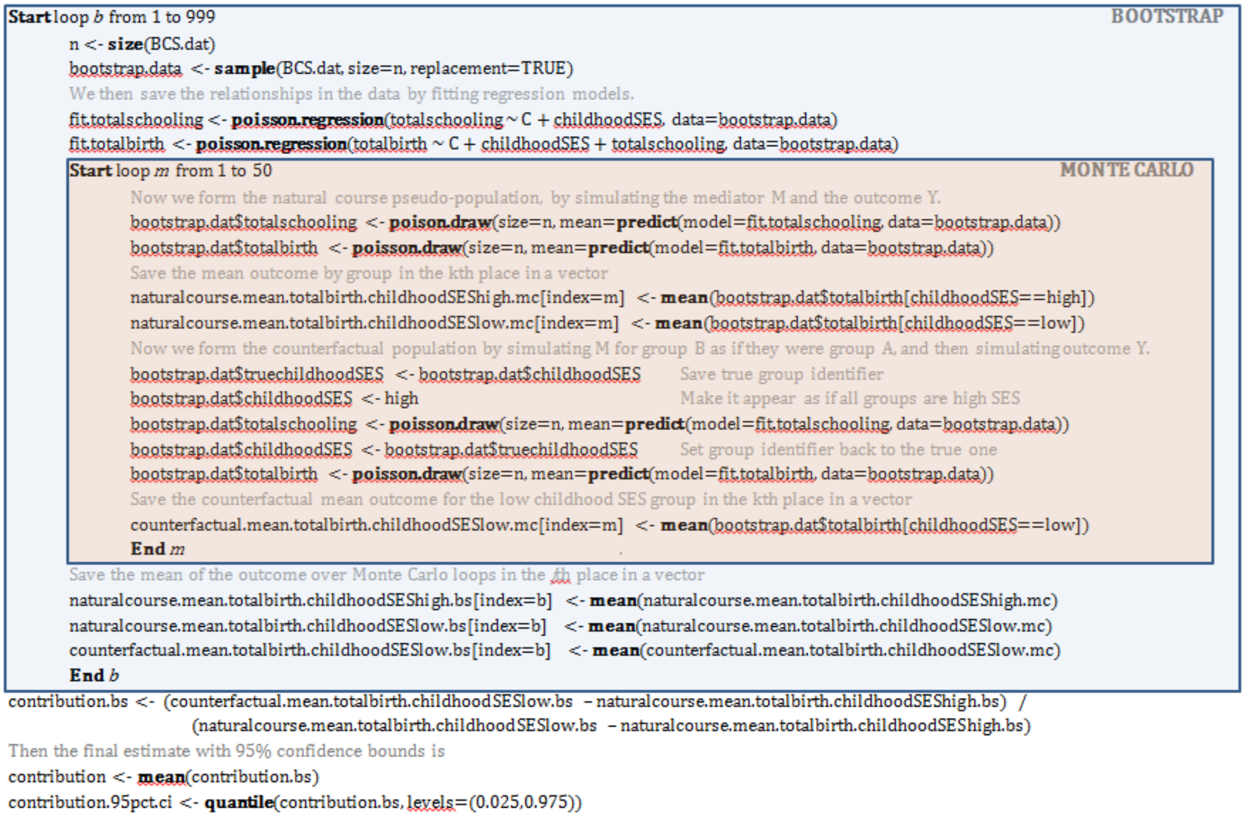
	Natural Course TFR	Counterfactual TFR	Natural Course Difference	Counterfactual Difference	Percent Mediated
High Childhood SES	1.45	1.45			
	(1.38, 1.51)	(1.38,1.51)			
Middle Childhood SES	1.55	1.49	0.10	0.04	58.2%
	(1.47,1.62)	(1.41,1.56)	(0.01,0.20)	(-0.05,0.14)	(19.7%,275.3%)
Low Childhood SES	1.72	1.63	0.27	0.19	30.9%
	(1.64,1.79)	(1.55,1.71)	(0.17,0.37)	(0.09,0.28)	(18.9%,52.2%)

Notes: Cohort TFR is only measured up to age 38 – sometimes called a partial TFR.

Table 3. Difference in period life expectancy at age 50 and the prevalence of ever regularly smoking cigarettes between men and women, Korean Longitudinal Study of Aging.

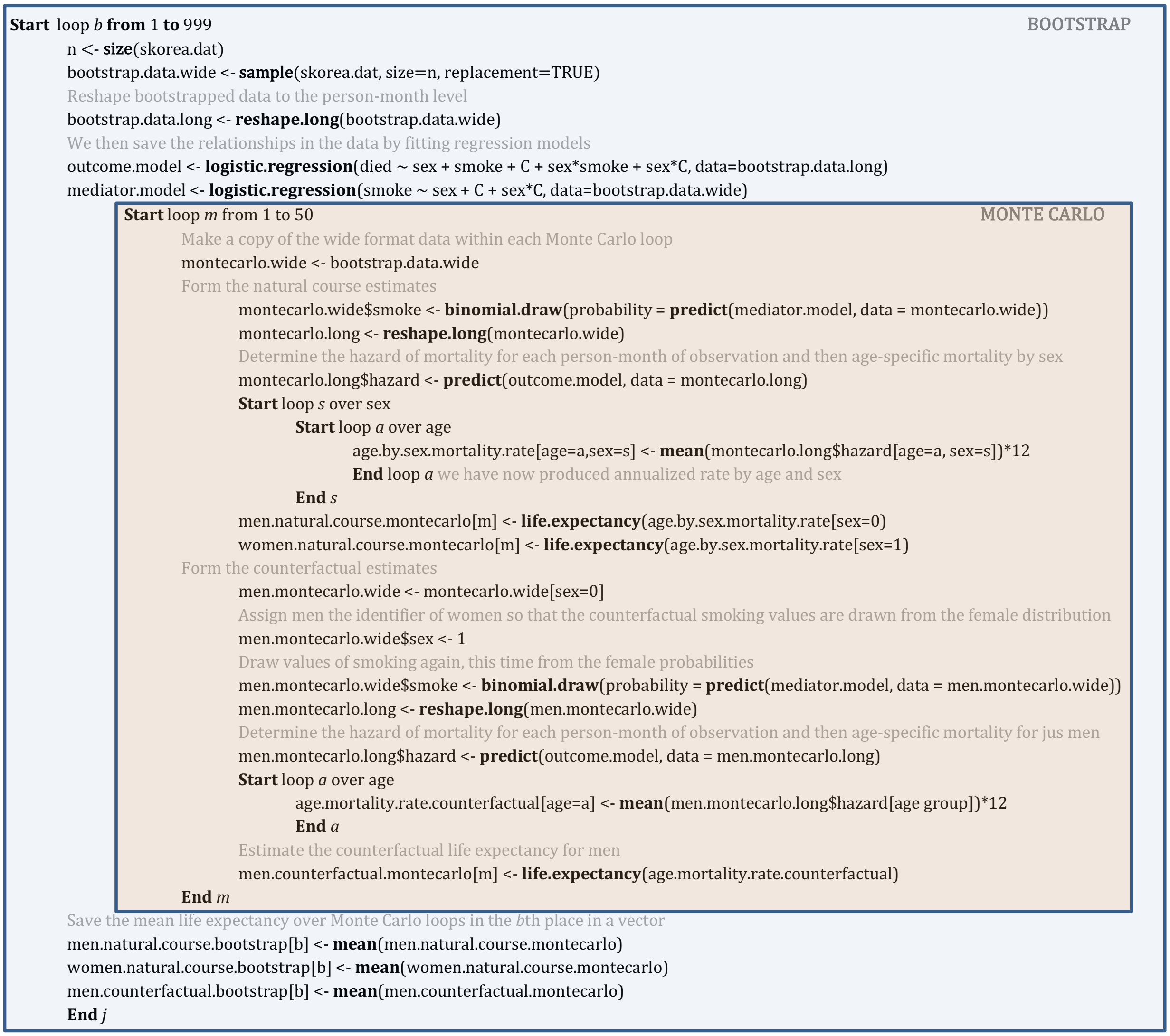
	Men	Women
Life expectancy at age 50	30.7	36.7
Prevalence of ever smoking	61.0%	4.3%

Notes: Data are from the 2006, 2008, 2010, and 2012 waves of the Korean Longitudinal Study of Aging. We estimated life expectancy at age 50 using standard period life table methods.



```
Start loop b from 1 to 999                                                                          BOOTSTRAP
        n <- size(BCS.dat)
        bootstrap.data  <- sample(BCS.dat, size=n, replacement=TRUE)
        We then save the relationships in the data by fitting regression models.
        fit.totalschooling <- poisson.regression(totalschooling ~ C + childhoodSES, data=bootstrap.data)
        fit.totalbirth  <- poisson.regression(totalbirth ~ C + childhoodSES + totalschooling, data=bootstrap.data)
        Start loop m from 1 to 50                                                               MONTE CARLO
                Now we form the natural course pseudo-population, by simulating the mediator M and the outcome Y.
                bootstrap.dat$totalschooling  <- poison.draw(size=n, mean=predict(model=fit.totalschooling, data=bootstrap.data))
                bootstrap.dat$totalbirth  <- poisson.draw(size=n, mean=predict(model=fit.totalbirth, data=bootstrap.data))
                Save the mean outcome by group in the kth place in a vector
                naturalcourse.mean.totalbirth.childhoodSEShigh.mc[index=m]  <- mean(bootstrap.dat$totalbirth[childhoodSES==high])
                naturalcourse.mean.totalbirth.childhoodSESlow.mc[index=m]  <- mean(bootstrap.dat$totalbirth[childhoodSES==low])
                Now we form the counterfactual population by simulating M for group B as if they were group A, and then simulating outcome Y.
                bootstrap.dat$truechildhoodSES  <- bootstrap.dat$childhoodSES       Save true group identifier
                bootstrap.dat$childhoodSES  <- high                                 Make it appear as if all groups are high SES
                bootstrap.dat$totalschooling  <- poisson.draw(size=n, mean=predict(model=fit.totalschooling, data=bootstrap.data))
                bootstrap.dat$childhoodSES  <- bootstrap.dat$truechildhoodSES       Set group identifier back to the true one
                bootstrap.dat$totalbirth  <- poisson.draw(size=n, mean=predict(model=fit.totalbirth, data=bootstrap.data))
                Save the counterfactual mean outcome for the low childhood SES group in the kth place in a vector
                counterfactual.mean.totalbirth.childhoodSESlow.mc[index=m]  <- mean(bootstrap.dat$totalbirth[childhoodSES==low])
                End m
        Save the mean of the outcome over Monte Carlo loops in the ith place in a vector
        naturalcourse.mean.totalbirth.childhoodSEShigh.bs[index=b]  <- mean(naturalcourse.mean.totalbirth.childhoodSEShigh.mc)
        naturalcourse.mean.totalbirth.childhoodSESlow.bs[index=b]  <- mean(naturalcourse.mean.totalbirth.childhoodSESlow.mc)
        counterfactual.mean.totalbirth.childhoodSESlow.bs[index=b]  <- mean(counterfactual.mean.totalbirth.childhoodSESlow.mc)
        End b
contribution.bs <-  (counterfactual.mean.totalbirth.childhoodSESlow.bs  – naturalcourse.mean.totalbirth.childhoodSEShigh.bs)  /
                    (naturalcourse.mean.totalbirth.childhoodSESlow.bs  – naturalcourse.mean.totalbirth.childhoodSEShigh.bs)
Then the final estimate with 95% confidence bounds is
contribution <- mean(contribution.bs)
contribution.95pct.ci <- quantile(contribution.bs, levels=(0.025,0.975))
```

Figure 1. Example bootstrap code for a Poisson mediator "totalschooling" (years in schooling) and Poisson outcome "totalbirth" (number of births between age 16 and 38). For the example in the figure, our summary measure is the mean number of children, our contrast of interest is the difference in mean fertility between women with low and high childhood SES, and for the counterfactual scenario we assign the education distribution of the high childhood SES group to the low childhood SES group. In the models, C represents covariates needed for exchangeability.



```
Start  loop b from 1 to 999                                                                    BOOTSTRAP
        n <- size(skorea.dat)
        bootstrap.data.wide <- sample(skorea.dat, size=n, replacement=TRUE)
        Reshape bootstrapped data to the person-month level
        bootstrap.data.long <- reshape.long(bootstrap.data.wide)
        We then save the relationships in the data by fitting regression models
        outcome.model <- logistic.regression(died ~ sex + smoke + C + sex*smoke + sex*C, data=bootstrap.data.long)
        mediator.model <- logistic.regression(smoke ~ sex + C + sex*C, data=bootstrap.data.wide)
                Start loop m from 1 to 50                                                      MONTE CARLO
                        Make a copy of the wide format data within each Monte Carlo loop
                        montecarlo.wide <- bootstrap.data.wide
                        Form the natural course estimates
                                montecarlo.wide$smoke <- binomial.draw(probability = predict(mediator.model, data = montecarlo.wide))
                                montecarlo.long <- reshape.long(montecarlo.wide)
                                Determine the hazard of mortality for each person-month of observation and then age-specific mortality by sex
                                montecarlo.long$hazard <- predict(outcome.model, data = montecarlo.long)
                                Start loop s over sex
                                        Start loop a over age
                                                age.by.sex.mortality.rate[age=a,sex=s] <- mean(montecarlo.long$hazard[age=a, sex=s])*12
                                                End loop a we have now produced annualized rate by age and sex
                                        End s
                                men.natural.course.montecarlo[m] <- life.expectancy(age.by.sex.mortality.rate[sex=0)
                                women.natural.course.montecarlo[m] <- life.expectancy(age.by.sex.mortality.rate[sex=1)
                        Form the counterfactual estimates
                                men.montecarlo.wide <- montecarlo.wide[sex=0]
                                Assign men the identifier of women so that the counterfactual smoking values are drawn from the female distribution
                                men.montecarlo.wide$sex <- 1
                                Draw values of smoking again, this time from the female probabilities
                                men.montecarlo.wide$smoke <- binomial.draw(probability = predict(mediator.model, data = men.montecarlo.wide))
                                men.montecarlo.long <- reshape.long(men.montecarlo.wide)
                                Determine the hazard of mortality for each person-month of observation and then age-specific mortality for jus men
                                men.montecarlo.long$hazard <- predict(outcome.model, data = men.montecarlo.long)
                                Start loop a over age
                                        age.mortality.rate.counterfactual[age=a] <- mean(men.montecarlo.long$hazard[age group])*12
                                        End a
                                Estimate the counterfactual life expectancy for men
                                men.counterfactual.montecarlo[m] <- life.expectancy(age.mortality.rate.counterfactual)
                        End m
        Save the mean life expectancy over Monte Carlo loops in the bth place in a vector
        men.natural.course.bootstrap[b] <- mean(men.natural.course.montecarlo)
        women.natural.course.bootstrap[b] <- mean(women.natural.course.montecarlo)
        men.counterfactual.bootstrap[b] <- mean(men.counterfactual.montecarlo)
        End j
```

Figure 2. Example bootstrap code for a binomial mediator "smoke" (ever smoker), binomial outcome "died" (death in a person-year). For the example in the figure, our summary measure is life expectancy at age 50, our contrast is the difference in life expectancy between men and women, and for the counterfactual scenario we assign men the smoking distribution of women. In the models, C represents covariates needed for exchangeability.

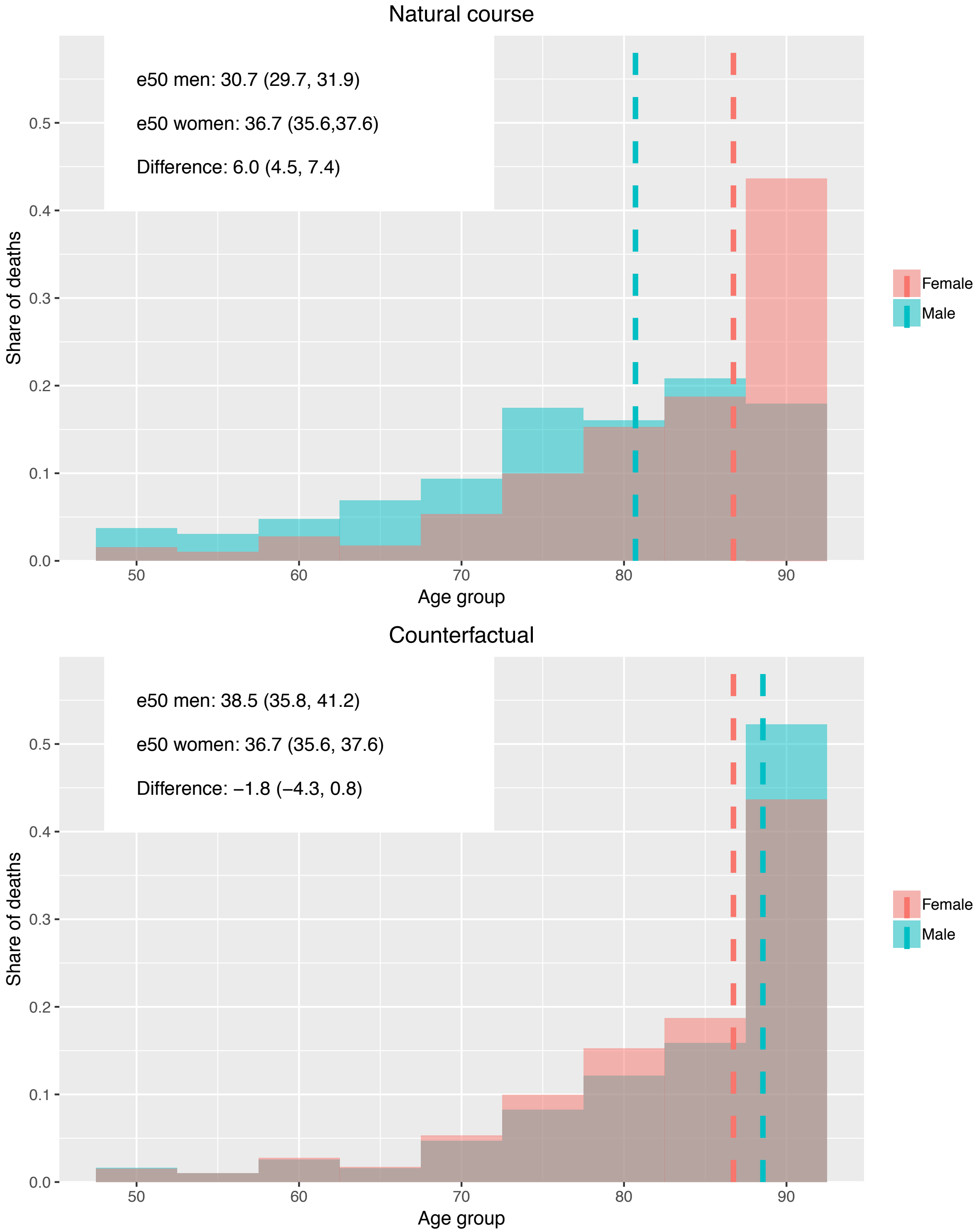


Figure 3. Natural course and counterfactual life table death distributions and estimates of the contribution of smoking to sex differences in life expectancy at 50, Korean Longitudinal Study of Aging.

Concept	Explanation
Grouping variable:	An indicator variable of the populations (or groups) to be compared.
Mediator:	A variable that potentially accounts for some of the differences between groups. In causal analysis, this variable should be a cause of the outcome of interest.
Counterfactual:	A hypothetical state of the world that may be different from its empirical state in some way. Standardized mortality rates are a common counterfactual used in demography: e.g. what would the mortality rate in the United States be if it had the age distribution of South Africa?
Causal effect:	In the Neyman-Rubin causal model, this is defined as the difference in an outcome when action A is present as compared with the outcome when action A is absent, all other things being equal (Neyman 1923; Rubin 1974).
Exchangeability:	The situation where two groups have the same distribution of covariates that affect the outcome of interest, with the exception of the grouping variable and mediators. This ensures that any differences in outcome between the two groups is due to differences in the determinant. In observational research, lack of exchangeability is often caused by confounding (or endogeneity).
Confounder:	Variables that affect both the mediator and the outcome of interest. Not adjusting for confounding variables violates the exchangeability assumption.
Equalizing:	Setting the distribution of a variable to be the same in two or more groups.
Pseudo-population:	A simulated population.
Natural course:	A pseudo-population that approximates the empirically observed data.
Monte Carlo estimation:	The process of approximating closed-form distributions by drawing simulated values repeatedly and averaging across the draws.

Box 1: Causal and *g*-computation terminology.

Appendix 1: Decomposition results when the mediator – outcome relationship is confounded

In this appendix, we show an example of how traditional demographic decompositions will provide incorrect contribution estimates when the relationship between the mediator and outcome of interest is confounded. For this example, we demonstrate this using a simple Kitagawa decomposition, but the same conclusion would hold for other decompositions that do not explicitly account for confounding.

Hypothetical Question

What is the contribution of systolic blood pressure (BP) to the difference in disability between two groups (group 1 and 2)?

Confounding

For this example, we will assume that systolic blood pressure and disability share the common cause of waist circumference (the confounder). This is not an unrealistic assumption, as many studies have shown that waist circumference affects BP and affects disability through other causes such as arthritis and diabetes.

Generating the data

We begin by drawing two separate distributions of waist circumference for groups 1 and 2 with a mean difference of 10 cm between groups:

```
# group size
g.size <- 1000000

# Waist circumference
waist.g1 <- rnorm(g.size, mean = 110, sd = 5)
waist.g2 <- rnorm(g.size, mean = 100, sd = 4)
```

We then create blood pressure values for each group using the following expressions:

```
# BP as a function of waist (using 0.35 as the relationship between waist and BP)
bp.g1 <- 130 + 0.35*waist.g1 + rnorm(g.size, mean = 0, sd = 5)
bp.g2 <- 120 + 0.35*waist.g2 + rnorm(g.size, mean = 0, sd = 6)
```

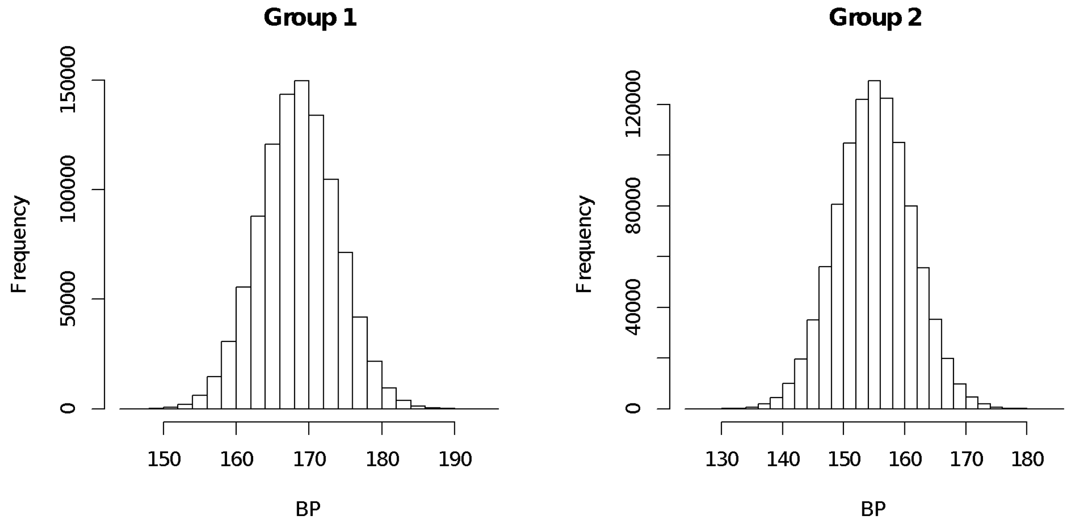
Finally we express the probability of disability as a logistic function of BP and waist circumference, and draw specific values of disability for individuals in both groups from these disability probabilities:

```
#Turn into data
toy <- data.frame(rbind(cbind(waist.g1, bp.g1, rep(1,g.size)),cbind(waist.g2, bp.g2, rep(2,g.size))))
colnames(toy) <- c("waist","bp","group")

# Disability as a function of both BP and waist
expit <- function(x) exp(x)/(1+exp(x))
toy$prob <- expit(-3+0.00499*toy$bp+0.00995*toy$waist)
toy$disability <- rbinom(2*g.size, size = 1, prob = toy$prob)
```

Generated Data

This results in the following data for BP:



With a prevalence of disability of 25.6% in group 1 and 22.6% in group 2.

Conducting the Decompositions

We first conduct a Kitagawa decomposition of disability between groups 1 and 2 by breaking up BP into bins of size 5 between 120 and 190 mmHg systolic BP. We then conduct two versions of the counterfactual decomposition, one where we did not control for confounding from waist circumference and one where we did. Since our decomposition requires specifying a counterfactual question, we estimate the answer to the question "How much smaller would the difference in the prevalence of disability between groups 1 and 2 be if they both had the same distribution?" To be consistent with the Kitagawa, we assign that distribution to be the average of the two distributions.

Results

Decomposition Table

	Kitagawa*	CFL Decomp no confounders	CFL Decomp with confounders
Contribution of BP	58%	58%	40%

*We show the contribution of difference in the distribution of BP term from the Kitagawa decomposition.

The results reveal that if we did not control for confounding for waist circumference, we would drastically overestimate the contribution of differences in the distribution of BP to the difference in disability between groups. This occurs because part of the contribution is driven by differences in the distribution of waist circumference and not BP. Secondly the results show that the in absence of controls for confounding the CFL decomp is equivalent to the Kitagawa, which is what we would expect based on results from JVW.

Could we have controlled for waist circumference in the Kitagawa Decomposition?

Technically there is nothing preventing us from controlling for waist circumference in the Kitagawa decomposition. However, doing so would require splitting waist circumference into bins, separately conducting the Kitagawa decomposition within the bins of waist circumference, and then reaggregating the results across bins. Even with just one confounder this rapidly creates large dimensionality issues, a problem that becomes even worse with multiple confounders.

Full R Code

Attached as appendix code.R.

Appendix 2: Other considerations regarding the application of the generalized counterfactual decomposition approach

Results from our approach are only valid if the underlying assumption of no unmeasured confounding of the mediator-outcome relationship is correct. Even with a large number of confounders, this is a strong assumption and thus the results need to be interpreted cautiously with consideration to the magnitude of bias that potential unmeasured confounders may introduce. Bias analyses, if adapted to our approach, may be a promising way to evaluate the causal validity of the decomposition estimates (Carnegie, Harada, and Hill 2016; VanderWeele and Arah 2011).

One conceptual issue that may arise is a lack of common support (also known as positivity) of the mediator distribution across groups. For example, suppose we are interested in equalizing the distribution of schooling between women with high and low childhood SES. If the low SES group has total education values of 6 to 9 and the high SES group has values ranging from 6 to 12, it cannot be determined from the data how the low education group would respond to having education values above 9. In such a case, one may be forced to assume that the relationship between total education values above 9 in the low SES group is the same as that of the high SES group or be willing to extrapolate the model estimates outside the range of observed data.

Finally, the selectivity of the sample must be considered in the interpretation of the results. In Example 3, we estimate the contribution of smoking to sex differences in mortality. To do so, we take a sample of individuals aged 50+ and estimate from them the relationship between smoking and mortality. When we assign the smoking distribution of women to men, we are not answering the question "what if men in general would have had the smoking distribution of women across their entire lives?", but instead "what if the men who survived to age 50 would have had the smoking distribution of women who survived to 50?". This distinction is important because some men may have died before age 50 and would have survived (and thereby potentially entered the sample) if they had the smoking distribution of women at ages before 50. This type of bias is known as selection or survivorship bias. If the interest is in the life course contribution of smoking, a sample covering all age groups should be taken.