medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Pathway-specific population attributable fractions

Maurice O'Connell[1] and John Ferguson [*1]

[1] *HRB Clinical Research Facility, National University of Ireland Galway*

[*] corresponding author

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Abstract

A population attributable fraction (PAF) represents the relative change in disease prevalence that one might expect if a particular exposure was absent from the population. Often, one might be interested in what percentage of this effect acts through particular pathways. For instance, the effect of excessive alcohol intake on stroke risk may be mediated by blood pressure, body mass index and several other intermediate risk factors. In this situation, attributable fractions for each mediating pathway of interest can be defined as the relative change in disease prevalence from disabling the effect of the exposure through that mediating pathway.

This quantity is related to, but distinct from the recently proposed metrics of direct and indirect PAF by Sjölander. In particular, while differing pathway-specific PAF will each usually be less than total PAF, they may sum over differing mediating pathways to more than total PAF, whereas direct and indirect PAF must sum to total PAF. Here, we present definitions, identifiability conditions and estimation approaches for pathway-specific attributable fractions. We illustrate results, and comparisons to indirect PAF using INTERSTROKE, a case-control study designed to quantify disease burden attributable to a number of known causal risk factors.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Pathway specific population attributable fractions

Introduction

Population attributable fractions (PAF) represent the relative change in disease prevalence that one might expect if a particular exposure was absent from the population. This metric was originally introduced by [1] in 1952 to estimate the percentage of lung cancer that would not have occurred under a counterfactual scenario that nobody smoked from the population. Since then, these family of metrics have become a standard way of measuring total disease burden attributable to a risk factor [2], and also to rank differing risk factors for prioritization as intervention targets [3].

Partitioning this overall disease burden into contributions from the known pathways through which the risk factor affects disease is also useful both in understanding pathogenic mechanisms and also when comparing interventions that may reduce disease. For instance, we might estimate that in a

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

hypothetical world, where dietary red-meat was completely substituted for plant based protein, the prevalence of heart disease might be reduced by 10%. How much of this reduction in disease burden is attributable to the pathway by which diet affects blood pressure? Here, we introduce the Pathway Specific population attributable fraction (PS-PAF) to help answer this question. In the preceding example, the PS-PAF can be informally understood as the relative change in disease prevalence in a hypothetical world where the distribution of blood pressure was altered to match the distribution expected under the afore-mentioned dietary substitution. Under certain assumptions (described later), the same quantity can be described more mechanistically as the proportion of disease that would be avoided from completely disabling the corresponding mediating pathway (here the pathway is diet $\rightarrow$ blood pressure $\rightarrow$ heart disease). PS-PAFs can be calculated for multiple mediating pathways for the same exposure. For example, the effect of the previous dietary substitution might be partially mediated by the effect on cholesterol as well as blood pressure; separate PS-PAFs could be compared for both pathways. In this case, the aim is not to provide a decomposition of the overall attributable fraction into the mediating pathways, and we argue later that such a decomposition would not be sensible. Just as differing attributable fractions for a set of risk factors typically sum to more than the joint attributable fraction [15], differing PS-PAFs corresponding to various mediating pathways between a particular risk factor and disease typically sum to more than the overall PAF for the risk factor. Rather than decomposing the total PAF, the aim instead is to fairly compare disease burden attributable to differing pathways and as a result gain insights into the dom-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

inant mechanisms by which the risk factor affects disease on a population level. These insights may in turn be useful in comparing possible interventions to prevent disease.

In the next Section, we will introduce notation and carefully describe PS-PAFs under various identifiability assumptions. We also will also compare and contrast PS-PAFs with the related concepts of indirect and direct PAF recently introduced by Sjölander [13]. Later, we illustrate results using data from INTERSTROKE, a case-control study designed to quantify disease burden attributable to a number of known causal risk factors for stroke. A further discussion section concludes the manuscript.

Methods

Potential outcome notation used for mediation analyses

We borrow notation from Vanderweele, [16], in defining nested counterfactuals. As is usual in the causal inference literature, random variables for observed variables will be denoted with unscripted notation, and potential outcomes will be denoted using subscripts. In all cases, we use upper case letters to denote random quantities, and lower-case to denote quantities that are fixed or intervened on. In particular, let C denote a vector of known baseline covariates not effected by the exposure, $A \in \{0, 1\}$ a binary exposure of interest and $M^1, ...M^K$ mediatos on separate causal pathways from A to Y, (note that each M^k could be binary, multi-category or continuous). Finally $Y \in \{0, 1\}$ is a binary disease outcome. Figure 1 below demonstrates

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

a multi-mediator scenario with 3 mediators, M^1, M^2 and M^3.

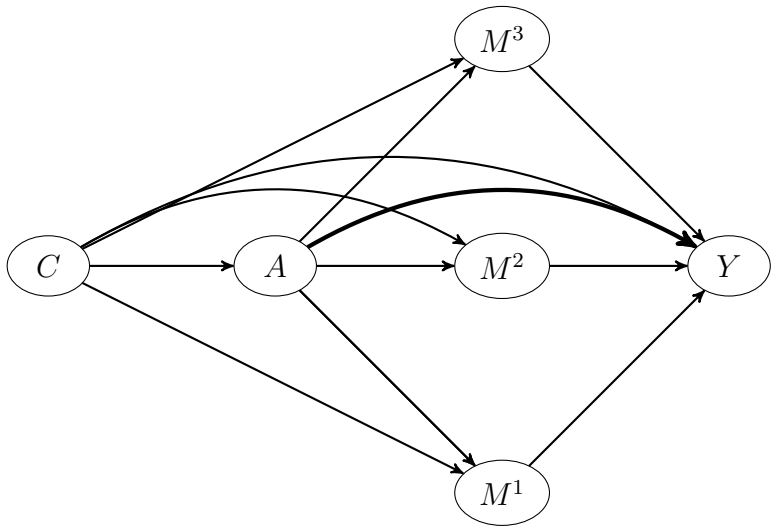

Figure 1: DAG showing causal structure linking exposure, mediators and outcome

The potential outcome setting exposure to a and the mediators to m^1,...m^K is denoted $Y_{a,m^1,\ldots,m^K}$. One can also define counterfactuals for each mediator k, assuming A is set to a as M_a^k. In addition, we will use the following abrieviated notation: $Y_a = Y_{a,M_a^1,\ldots,M_a^k}$, $Y_{A,m^j} = Y_{A,m^1,m^2,\ldots,m^K}$ where $m^k = M_A^k$ for $k \neq j$, and $Y_{0,M^j} = Y_{0,m^1,m^2,\ldots,m^K}$, where $m^j = M^j$ and $m^k = M_0^k$ for $k \neq j$.

As is usual with causal inference using the potential outcomes framework, we make Stable Unit Treated Value Assumptions (SUTVA) [18], which implies that the relationship between potential and observed outcomes satisfy 'consistency', or that $Y = Y_{a,M_a}$ when $A = a$. In addition, mediation analysis requires some technical conditional independence assumptions which we will list as needed in the following sections.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

The population attributable fraction

Despite being an intrinsically causal idea, attributable fractions were originally defined via conditional probabilities in a non-causal framework, which has contributed to confusion regarding what they really purport to measure. Here, we will instead use a definiton based on potential outcomes which is now becoming more prominent [9, 8].

$$PAF_{total} = \frac{P(Y=1) - P(Y_0=1)}{P(Y=1)}. \tag{1}$$

Thinking of Y_0 as the potential outcome for a random individual in a population where no-one was exposed to the risk factor, (1) can be directly interpreted as the relative change in disease prevalence if that risk factor was absent from the population.

The Pathway Specific population attributable fraction

For a binary risk factor $A \in \{0, 1\}$, the PS-PAF for the mediating pathway $A->M^j->Y$ is defined as

$$PAF_{A->M^j->Y} = \frac{P(Y=1) - P(Y_{A,G^j_{0|C}}=1)}{P(Y=1)} \tag{2}$$

where the random variable $G^j_{0|C}$ is a random draw from the conditional distribution of M^j_0 given an individual's covariates, C. Marginally, G^j_0 (simulated via first randomly sampling an individual from the population and using their covariates, C, to generate $G^j_{0|C}$) is a draw from the population distribution of M^j_0, that is the distribution that M^j would have in a hypo-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

thetical world where the risk factor was eliminated. Noting this, the PS-PAF can then be interpreted in terms of a population intervention on M^j, where its distribution is shifted to the distribution that would be observed in a population where the risk factor A was eliminated, with other quantities in the population (covariates C, exposure A and other mediators $M^k, k \neq j$) remaining unchanged. This idea of randomly assigned interventions has previously been introduced to estimate versions of natural direct and indirect effects (sometimes termed interventional direct and indirect effects) in the presence of exposure-induced confounding [**?**]. Comparing PS-PAFs for the various known mediating pathways can help in determining the predominant pathways by which the risk factor is effecting disease in the population. One might also want to compare disease burden attributable to pathways that are either unknown (or involve unobserved mediators). To do this, we will adapt the concept of Direct PAF proposed by Sjölander [13] as follows:

$$PAF_{A->Y} = \frac{P(Y=1) - P(Y_{0,M^1,\ldots,M^K}=1)}{P(Y=1)} \tag{3}$$

We will refer to equation (3) as the PS-PAF for the direct pathway $A->Y$. (3) measures the disease burden attributable to direct pathogenic pathways from A to Y, that is any pathway from A to Y (that may or may not be known) excluding the set of mediating pathways under consideration. The definition in [13] is very similar, except that it refers to the direct pathway with reference to a single mediating pathway, j, and will change dependent on that mediating pathway as follows:

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

$$PAF_{direct,M^j} = \frac{P(Y=1) - P(Y_{0,M^j}=1)}{P(Y=1)} \tag{4}$$

As alluded to in the previous section, $Y_{0,M^j} = Y_{0,m^1,m^2,...,m^K}$, where $m^j = M^j$ and $m^k = M_0^k$ for $k \neq j$, and is not necessarily equal to $Y_{0,M^1,...,M^K}$.

Identifiability conditions

Idetifiability conditions are needed to estimate (2) even when values on the exposure, mediators, observed covariates and outcome are available. In particular, conditions 1. and 2. below are necessary to identify PS-PAFs for each mediator j of interest.

1. $M_0^j \perp\!\!\!\perp A|C$ (informally, associations between the risk factor and mediator have causal interpretations within strata of covariates)

2. $Y_{a,m^j} \perp\!\!\!\perp M^j|A,C$ (informally, associations between the mediator and outcome have causal interpretations within joint strata of risk factor and covariates)

An additional condition is necessary to identify the $PAF_{A->Y}$

3. $Y_{a,M^1,...,M^K} \perp\!\!\!\perp A|M^1,...,M^K,C$, (informally, associations between the risk factor and outcome have causal interpretation within joint strata of mediators $(M^1,...,M^K)$ and covariates)

Under identifiability conditions 1, 2 and 3 we show in the supplementary material that:

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

$$PAF_{A->M^j->Y} = \frac{P(Y=1) - E_{A,C}(E_{M^j|A=0,C}(P(Y=1|A,C,M^j)))}{P(Y=1)} \quad (5)$$

and

$$PAF_{A->Y} = \frac{P(Y=1) - E_{C,M^1,...,M^K}(P(Y=1|A=0,C,M^1,...,M^K))}{P(Y=1)} \quad (6)$$

where $E_{A,C}(g(A,C)) = \int_{a,c} g(a,c)dF_{A,C}(a,c)$ and $E_{C,M^1,...,M^K} g(C,M^1,..,M^K) = \int_{c,m^1,...,m^K} g(a,m^1...,m^K)dF_{A,M^1,...,M^K}(a,m^1,...,m^K)$ represent expectations integrated over the marginal distributions $F_{A,C}$ and $F_{A,M^1,...,M^K}$ of the subscripted variables and $E_{M^j|A=0,c}(g(M^j)) = \int_{m^j} g(m^j)dF_{M^j|A=0,C=c}(m^j)$ is an expectation integrated according to the conditional distribution of M^j given $A=0$ and $C=c$, $F_{M^j|A=0,C=c}$. Note that condition 2. implies that there is no post treatment confounding of the mediator outcome relationship; that is there is no variable L that is affected by the risk factor A such that L is a joint cause of M^j and Y. If such a variable L exists, condition 2. would then become $Y_{a,m^j} \perp\!\!\!\perp M^j|A,C,L$, and the pathway specific PAF would change to

$$PAF_{A->M^j->Y} = \frac{P(Y=1) - E_{A,C,L}(E_{M^j|A=0,C,L}(P(Y=1|A,C,L,M^j))}{P(Y=1)}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Mechanistic pathway specific PAF and disabling pathways

Under an extra identifiability condition, the pathway specific PAF can be also expressed in a mechanistic form where the mediator assignment to an individual (within the hypothetical population where the distribution of the mediator is altered) is the mediator that would result for that individual under no exposure to the risk factor:

$$PAF_{A->M^j->Y} = \frac{P(Y=1) - P(Y_{A,M_0^j}=1)}{P(Y=1)} \tag{7}$$

This final condition is:

4. $Y_{a,m^j} \perp\!\!\!\perp M_0^j | A = a, C$.

This condition is less intuitive than the conditions 1., 2. and 3. described in the preceding section and involve consideration of cross-world counterfactuals (that is if $a = 1$, Y_{a,m^j} and M_0^j would never be observed on the same individual). However, as shown in the supplementary material, this condition does hold in a non-parametric structural equations model. Under this mechanistic interpretation, the pathway specific PAF can be thought of as the relative change in disease burden in a hypothetical population where the mediated pathway $A->M^j->Y$ was disabled. For example, in a simple setting where there is only a single known mediator, M, there are 2 potential pathways by which the risk factor effects disease, represented by the pathways $A->M->Y$ and $A->Y$ in Figure 2(a). The total PAF (which

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

corresponds to a population where a binary risk factor was eliminated corresponds to the disabling both pathways, that is a comparison of disease risk in the populations with causal graphs show in the left hand and right hand panes of Figue 2(a). In contrast, direct PAF involves a comparison of disease risk in the current population with the hypothetical population where the direct pathway is disabled (Figure 2(b)) whereas the pathway specific PAF represents a comparison of the current population with a hypothetical population where the pathway $A->M->Y$ is disabled (Figure 2c). Note that in each case, current disease risk is compared to disease risk in some hypothetical population where a pathway has been disabled.

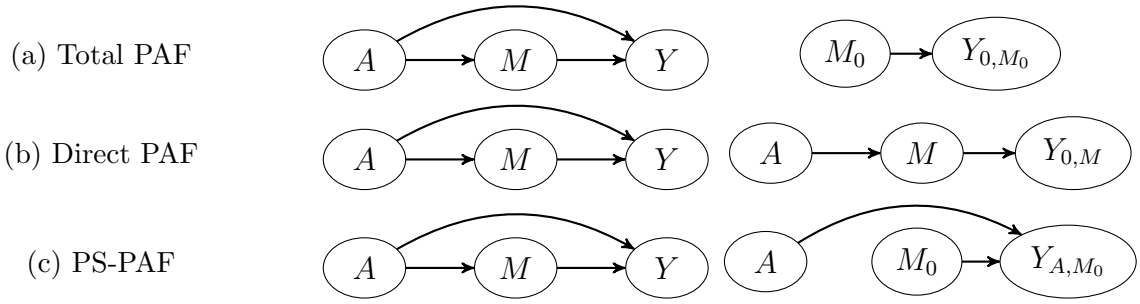

Figure 2: Illustration of total, direct and pathway specific PAF as a comparison between disease risk in current population (graph on LHS) and counterfactual disease risk in a population where particular pathways through which the risk factor effects disease are disabled (graph on RHS). A situation with a single mediated pathway $A->M->Y$ is illlustrated. Sub-figure (a) Total PAF compares observed disease risk $P(Y=1)$ vs disease risk $P(Y_{0,M_0}=1)$ in a hypothetical population where both direct and mediated pathways are disabled. Sub-figure (b) Direct PAF compares observed disease risk vs disease risk $P(Y_{0,M}=1)$ in a hypothetical population where the direct pathway is disabled. Sub-figure (c) Pathway specific PAF compares observed disease risk Y vs disease risk $P(Y_{A,M_0}=1)$ in a hypothetical population where the mediated pathway $A->M->Y$ is disabled

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Pathway Specific PAF vs indirect PAF

In addition to proposing the direct PAF (PAF_{direct,M^j}, defined in equation (4)) [13] introduced the concept of indirect PAF, defined as:

$$PAF_{indirect,M^j} = PAF_{total} - PAF_{direct,M^j} \tag{8}$$

defined in the context of a single mediating pathway, so that the direct and indirect attributable fractions sum to the total PAF. Rather than comparing disease risk in the current and hypothetical populations, (8) implicitly compares disease risk in two hypothetical popualtions: one where the direct pathway $A->Y$ has been disabled, with a second where the direct and mediated pathway are both disabled (see Figure 3). Indirect PAF will usually be smaller than the corresponding pathway specific PAF in a one mediator situation, as it is likely that some disease cases in the current population (which are exposed to the risk factor) might be equally well prevented by either eliminating the effect of the direct pathway (that is perhaps $Y_{1,M_1} = 1$, but $Y_{0,M_1} = 0$, or eliminating the mediated pathway (that is perhaps $Y_{1,M_1} = 1$, but $Y_{1,M_0} = 0$). The prevention of such disease cases would contribute to the pathway specific PAF, but not to the indirect PAF.

Estimation

In cohort and cross-sectional studies (5) and (6) can be estimated without bias from empirically estimated conditional distributions, given the iden-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 3: Indirect PAF compares disease risk $P(Y_{0,M} = 1)$ in a hypothetical population with the direct pathway disabled with the disease risk $P(Y_{0,M_0} = 1)$ in a hypothetical population with both direct and mediated pathways have been disabled.

tifiability conditions discussed earlier. For case control studies, we use a simple-reweighting trick which assumes that the prevalence of disease, π, is known and the sampled disease cases and controls are randomly selected from their respective populations. We assume for simplicity that the case:control matching ratio is 1:r, for some $r \geq 1$. Under these assumptions, the components of (5) and (6) can be found using a re-weighted dataset where cases are assigned weights $w_i = 1$, and controls are assigned weights $w_i = (1/\pi - 1)/r$. (Note that under the assumptions that prevalence is known and the cases and controls are randomly selected from their source populations, this reweighted sample is a representative sample from the source population; that is means of any variable in this reweighted population agree with the source population, and statistical estimation is consistent). Suppose then, that the researcher specifys and estimates correct models for $P(Y = 1|A, C, M^1, ..., M^K)$ and a correct conditional model for each M^k, conditioned on A and C (perhaps in the re-weighted population if the original data is from a case-control study). To avoid numerical integration in the continuous mediator case, and simplify the algorithm, we propose that for mediator j, $E_{A,C}(E_{M^j|A=0}(P(Y = 1|A, C, M^j)))$ is estimated via the following algorithm. In the algorithm, we set $j = 1$ for simplicity of notation, however an algorithm with obvious

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

modification can be used to estimate the pathway specific PAF for other mediators $j \leq K$. The algorithm assumes that the mediators $M^1, ..., M^K$ are conditionally independent given A and C, which implies for instance that $E_{A,C}(E_{M^1|A=0,C}(P(Y = 1|A, C, M^1))$=$E_{A,C,M^2,...,M^K}(E_{M^1|A=0,C}(P(Y = 1|A, C, M^1, ..., M^K)))$. With this assumption, only a single model for the outcome: $P(Y = 1|A, C, M^1, ..., M^K)$ that depends on all mediators M^1,...,M^K needs to be specified. If the user prefers not to assume this conditional independence condition, the algorithm could be modifed by fitting K separate models for Y each using a single mediator; that is the jth model would regress Y on A, C and M^j to estimate $P(Y = 1|A, C, M^j)$.

1. Choose a number of simulation iterates, S
2. for(j in 1:S){
   (a) For each individual in the data, i, with exposure, A_i, covariate vector C_i and mediators $M_i^1, ..., M_i^K$ simulate M^{1*} from the estimated conditional distribution of M^1 given $A_i = 0$ and C_i
   (b) Calculate $\hat{P}(Y = 1|A_i, C_i, M^{1*}, M_i^2, ...M_i^K)$ for each i using the fitted statistical model
   (c) Calculate $E_j = \sum_{i \leq N} \widehat{P}(Y = 1|A_i, C_i, M^{1*}, M_i^2, ..., M_i^K)/N$ for cohort/cross sectional designs; calculate $E_j = \sum_{i \leq N} w_i \widehat{P}(Y = 1|A_i, C_i, M^{1*}, M_i^K, ..., M_i^K)/\sum_{i \leq N} w_i$ for case control designs.

   }

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

3. $E_{A,C}(E_{M^1|A=0,C}(\widehat{P(Y} = 1|A, C, M^j))) = \sum_{j \leq S} E_j/S$

To estimate direct PAF, $E_{C,M^1,...,M^K}(P(Y = 1|A = 0, C, M^1, ...M^K)))$ can be estimated more simply by estimating $P(Y = 1|A = 0, C_i, M_i^1, ..., M_i^K)$ for each individual, and averaging over individuals in the data, taking care to incorporate weighting under a case control design. As an alternative, a double robust estimator for $E(P(Y = 1|A = 0, C, M^1, ..., M^K))$ can be derived using the same approaches Sjölander describes in [13].

Data Example

INTERSTROKE, [20], is a large international case control study designed to quantify the contribution of established risk factors to stroke prevalence at a global level. Here we investigate the possible mediating effects of physical activity on stroke through waist hip ratio, apolipoproteins B to A1 ratio and diagnosis of high blood pressure. We treat waist hip ratio and apolioproteins as continuous variables, whereas diagnosed high blood pressure is binary. Covariates and assumed mediators are as assumed in the causal structure shown by Figure 4. To estimate Sjölander's direct and indirect attributable fractions, and the pathway specific attributable fractions described above, we fit a main-effects logistic regression predicting stroke status as a function of age,sex,region,eduction diet,physo-social stress factors,smoking status, alcohol use, physical activity, waist hip ratio, apoB/apoA ratio and clinically diagnosed high blood pressure, with the terms for wasit hip ratio and apoB/apoA ratio entering as 5-degree of freedom natural cubic splines. In this regression, stroke controls were upweighted by a factor of 284 to re-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

flect a yearly stroke incidence of first stroke of 0.0035 or 3.5 strokes per 1000 individuals per year, estimated via data from the global burden of disease [21]. To apply the algorithm in the previous section, we also need to simulate from the predicted distribution for the mediators, ApoB/ApoA ratio, waist hip ratio and clincially diagnosed hypertension (conditioned on values for age, sex, education, region, physical activity, diet-score, stress, smoking and alcohol). To do this, we resample residuals from a fitted linear model for ApoB/ApoA ratio and waist hip ratio and add these resampled residuals to the fitted values, whereas for hypertension we simply draw Bernoulli variables with probabilities with probabilities according to the fitted logistic models.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Pathway	**total PAF**	$PAF_{A->Y}$	PAF_{Direct,M^j}	$PAF_{Indirect,M^j}$	$PAF_{A->M^j=>Y}$
PHYS->HBP->STROKE	0.40 (0.34, 0.47)	0.35 (0.27, 0.42)	0.39 (0.33, 0.46)	0.013 (-0.014, 0.039)	0.041 (0.014, 0.068)
PHYS->WHR->STROKE	0.40 (0.34, 0.47)	0.35 (0.27, 0.42)	0.40 (0.33, 0.47)	0.005 (-0.016, 0.026)	0.032 (0.007, 0.058)
PHYS->APOB->STROKE	0.40 (0.34, 0.47)	0.35 (0.274, 0.42)	0.39 (0.32, 0.45)	0.018 (-0.005, 0.041)	0.045 (0.016, 0.073)

Table 1: Results from estimating the pathway specific, direct and indirect PAFs on the INTERSTROKE dataset using S=100 mediator simulations. Estimated 95% confidence intervals (estimate $\pm 1.96 *$standard error) are shown in brackets. Standard errors were estimated using 100 bootstrap iterations, with one bootstrap iteration involving independent resamples from the sample of controls and cases. PAF estimates are rounded to 2 significant digits

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Note that while there is only one pathway specific PAF from A to Y, Sjolander's definition of direct PAF is with reference to a single mediating pathway and will change depending on that pathway. For instance, in this example the direct PAF with reference to high blood pressure essentially pools together all other pathways not going through high-blood pressure. In this example, the mediating pathways and direct pathway due to lack of exercise are likely to increase the probability of stroke, and one would expect PS-PAFs to be somewhat larger than the corresponding indirect PAFs. This pattern is seen in Table 1; the estimated pathway specific PAFs when expressed as percentages range from 3.2%-4.5%, whereas the indirect PAFs range from 0.5% to 1.8%. On the other hand, $PAF_{A->Y}$ (defined in equation (3)) is 34.3%. According to this analysis, while population disease burden for stroke attributable to physical activity may partially depend on the mediating pathways through blood pressure, waist hip ratio and lipids, it depends heavily on other (unknown) mechanisms. As with any causal analysis, these tentative conclusions depend jointly on correct modeling of conditional probability distributions and on the validity of the causal identifiability assumptions that we listed earlier. Sensitivity analysis [16] could be performed to guage the biases in these results that might be expected due to violations of these conditions. However, the differences between $PAF_{A->Y}$ and pathway specific mediated PAFs observed here are so large that it would be hard to completely attribute these differences to incorrect modeling or an incomplete causal model.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Discussion

In this paper we have introduced pathway specific attributable fractions as a metric to measure the disease burden attributable to particular exposure mediator pathways. The idea is related to the recently proposed ideas of direct and indirect PAF by Sjölander, but more suitable to multiple mediators [13]. Some differences between the two metrics are summarized in Table 2. In a one mediator situation, one way to understand the distinctions between these concepts is in terms of modified sequential attributable fractions [15], that is attributable fractions that are constructed from disabling pathways in a particular order (as demonstrated in Figures 2 and 3) with the order in which pathways are disabled differing for PS-PAFs and indirect PAFs. In more detail, a PS-PAF can be interpreted as the relative change in disease burden from disabling a particular mediating pathway. The corresponding indirect PAF is also associated with disabling that same mediating pathway, but this time subsequent to disabling the direct pathway. Since the effect of disabling both the direct and mediating pathways is equivalent to the effect of eliminating the risk factor, this effectively forces the additivity property that total PAF is the sum of direct and indirect PAF. Note that in general sequential attributable fractions are constructed to sum to some well-defined overall PAF but usually the sequence corresponds the hypothetical elimination of each of a group of risk factors in some order, rather than disabling pathways for a particular risk factor in some order.

While this additivity property at first seems appealing, it perhaps is unnatural in the context of attributable fractions, where it is well recognized

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

that the PAF for differing risk factors may sum to more than the joint PAF and sometimes to more than 1 [22]. The sufficient/component cause framework [23] gives a simple but enlightening explanation for this phenomenon. For particular individuals, a certain collection of risk factors (perhaps diet, stress and tobacco usage) might collectively lead to disease at a particular point in time, but the disease may not have occurred at that time if any of the risk factors were not present. This implies that pathway specific PAFs will tend to be larger than indirect PAFs as illustrated in this manuscript, if we view the direct and indirect pathways as independent disease-causing mechanisms.

Whereas attributable fractions can measure the total disease burden associated with a risk factor, there are less useful to measure the real-world impact of a public health intervention on that risk factor since even successful health interventions usually only partially eliminate a risk factor and in addition cannot alter prior history to a risk factor when cumulative exposure might also impact disease. As an example, rather than considering a hypothetical population where smoking is elimated, a realistic population level intervention (such as increasing the tax on cigarettes) may result in a 5% decrease in the number of cigarettes consumed rather than total elimination of smoking. Impact fractions are generalized versions of attributable fractions that measure the reduction in disease prevalence associated with such a population intervention. The ideas described here can easily be adapted to define and estimate pathway specific impact fractions for such real world interventions which may characterize the dominant mechanisms by which the intervention affects disease burden. For example, the pathway specific

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

impact fraction for the pathway $A->M^j->Y$ could be defined by letting $G^j_{|C}$ represent a random variable having the population distribution of the mediator M^j under the proposed intervention (simulated again conditional on an individuals covariate vector C) and replacing $G^j_{0|C}$ with $G^j_{|C}$ in equation (2).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

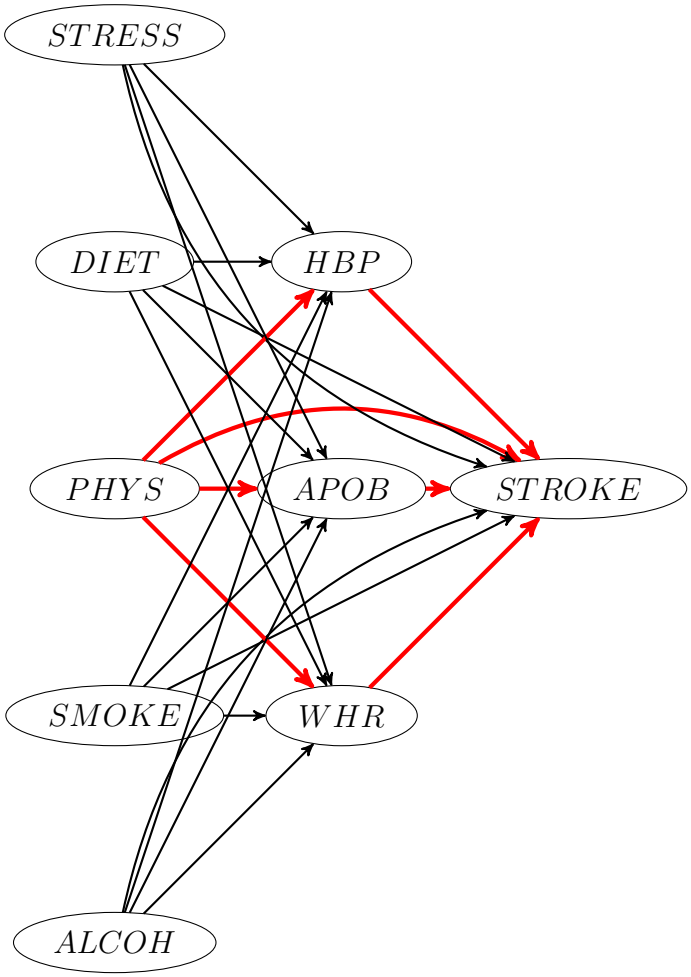

Figure 4: DAG showing assumed causal structure for risk factors in INTERSTROKE. The direct and mediating pathways associated with physical activity are highlighted in bold red. Age, sex and geographic region are confounders for the risk/factor disease relationship for all listed factors on the figure (these are omitted for display purposes)

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

	Indirect PAF	**Pathway specific PAF**
Number of mediators	Focused on one mediator, as the difference between total PAF and direct PAF	Naturally Focused on several mediators
Interpretation of direct PAF	Defined in context of a single mediator as what the PAF would be if the exposure didn't affect the distribution of the mediator	Represents the PAF attributable to none of the measured mediators. Can be regarded as a PS-PAF for all unobserved mediating pathways
Interpretation of mediated PAF	The extra attributable risk that is left over from the total PAF after subtracting the direct PAF	The disease burden associated with a hypothetical intervention setting the distribution of the mediator to the distribution expected under elimination of the risk factor
Addivity	Indirect PAF+direct PAF =overall PAF by definition	Additivity not present by default.
Relative size of mediated PAF	Might be surprisingly small due to interpretation as seqential PAF corresponding to disabling the indirect pathway sequentially after the direct pathway already disabled	Usually larger than indirect PAF when disease risk factors are positively correlated

Table 2: Comparison of PS-PAF with indirect and direct PAF

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

References

[1] L. ML, "The occurrence of lung cancer in man.," *Acta-Unio Internationalis Contra Cancrum*, vol. 9, no. 3, pp. 531–541, 1953.

[2] C. Poole, "A history of the population attributable fraction and related measures," *Annals of epidemiology*, vol. 25, no. 3, pp. 147–154, 2015.

[3] J. Ferguson, N. O'Leary, F. Maturo, S. Yusuf, and M. O'Donnell, "Graphical comparisons of relative disease burden across multiple risk factors," *BMC medical research methodology*, vol. 19, no. 1, p. 186, 2019.

[4] P. Cole and B. MacMahon, "Attributable risk percent in case-control studies.," *British journal of preventive & social medicine*, vol. 25, no. 4, p. 242, 1971.

[5] O. S. Miettinen, "Proportion of disease caused or prevented by a given exposure, trait or intervention," *American journal of epidemiology*, vol. 99, no. 5, pp. 325–332, 1974.

[6] P. Bruzzi, S. B. Green, D. P. Byar, L. A. Brinton, and C. Schairer, "Estimating the population attributable risk for multiple risk factors using case-control data," *American journal of epidemiology*, vol. 122, no. 5, pp. 904–914, 1985.

[7] O. Gefeller, "An annotated bibliography on the attributable risk," *Biometrical Journal*, vol. 34, no. 8, pp. 1007–1012, 1992.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

[8] E. Suzuki, E. Yamamoto, and T. Tsuda, "On the relations between excess fraction, attributable fraction, and etiologic fraction," *American journal of epidemiology*, vol. 175, no. 6, pp. 567–575, 2012.

[9] M. A. Hernán and J. M. Robins, "Causal inference: what if," *Boca Raton: Chapman & Hill/CRC*, vol. 2020, 2020.

[10] R. M. Baron and D. A. Kenny, "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.," *Journal of personality and social psychology*, vol. 51, no. 6, p. 1173, 1986.

[11] J. M. Robins and S. Greenland, "Identifiability and exchangeability for direct and indirect effects," *Epidemiology*, pp. 143–155, 1992.

[12] J. Pearl, "Direct and indirect effects," *arXiv preprint arXiv:1301.2300*, 2001.

[13] A. Sjölander, "Mediation analysis with attributable fractions," *Epidemiologic Methods*, vol. 7, no. 1, 2018.

[14] G. E. Eide and O. Gefeller, "Sequential and average attributable fractions as aids in the selection of preventive strategies," *Journal of clinical epidemiology*, vol. 48, no. 5, pp. 645–655, 1995.

[15] J. Ferguson, M. O'Connell, and M. O'Donnell, "Revisiting sequential attributable fractions," *Archives of Public Health*, vol. 78, p. 67, July 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

[16] T. VanderWeele, *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.

[17] J. Pearl, M. Glymour, and N. P. Jewell, *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[18] D. B. Rubin, "Randomization analysis of experimental data: The fisher randomization test comment," *Journal of the American Statistical Association*, vol. 75, no. 371, pp. 591–593, 1980.

[19] J. Ferguson, A. Alvarez-Iglesias, J. Newell, J. Hinde, and M. O'Donnell, "Estimating average attributable fractions with confidence intervals for cohort and case–control studies," *Statistical methods in medical research*, vol. 27, no. 4, pp. 1141–1152, 2018.

[20] M. J. O'Donnell, S. L. Chin, S. Rangarajan, D. Xavier, L. Liu, H. Zhang, P. Rao-Melacini, X. Zhang, P. Pais, S. Agapay, *et al.*, "Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (interstroke): a case-control study," *The lancet*, vol. 388, no. 10046, pp. 761–775, 2016.

[21] "Gbd compare data visualization, seattle, wa: Ihme, university of washington, 2018. available from http://vizhub.healthdata.org/gbd-compare,"

[22] A. K. Rowe, K. E. Powell, and W. D. Flanders, "Why population attributable fractions can sum to more than one," *American journal of preventive medicine*, vol. 26, no. 3, pp. 243–249, 2004.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.15.20212845; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

[23] K. J. Rothman, "Causes," *American journal of epidemiology*, vol. 104, no. 6, pp. 587–592, 1976.

[24] J. Pearl, *Chapter in Book: C. Berzuini, P. Dawid, and L. Bernardinelli (Eds.), Causality: Statistical Perspectives and Applications*, ch. 12. The Mediation Formula: A guide to the assessment of causal pathways in nonlinear models, pp. pp. 151–179. 2012.