Published in final edited form as:

Eur J Epidemiol. 2013 October ; 28(10): . doi:10.1007/s10654-013-9843-6.

Reconsidering the denominator of the attributable proportion for interaction

Tyler J. VanderWeele

Harvard School of Public Health, Boston, MA, USA

Tyler J. VanderWeele: tvanderw@hsph.harvard.edu

A number of measures have been proposed to assess interaction on the additive or risk difference scale between the effects of two exposures. Assessing such additive interaction is desirable because it allows one to assess whether it would be better to target an intervention to certain subgroups if resources are limited [1–3]. Additive interaction also allows one to assess evidence for mechanistic interaction within the sufficient cause or potential outcomes frameworks [2–5]. One of the measures that has been proposed to assess additive interaction is what is sometimes referred to as the attributable proportion [6]. We will consider the properties of two different definitions of this attributable proportion measure.

We let D denote a binary outcome, X_1 and X_2 two binary exposures, and $p_{x_1x_2} = P(D = 1|X_1 = x_1, X_2 = x_2)$ the probability of the outcome when $X_1 = x_1, X_2 = x_2$. We will assume that analysis is conducted within strata of covariates that suffice to adjust for confounding of the effects of both exposures. Additive interaction on the risk difference scale is defined as $(p_{11} - p_{10} - p_{01} + p_{00})$, sometimes also referred to as the interaction contrast [2]. We can rewrite this additive interaction measure as $(p_{11} - p_{10} - p_{01} + p_{00}) = (p_{11} - p_{00}) - \{(p_{10} - p_{00}) + (p_{01} - p_{00})\}$ to see that the measure essentially captures the extent to which the effect, on the risk difference scale, of both exposures combined exceeds the sum of the effects of the first exposure considered alone and the second exposure considered alone.

This measure of additive interaction on the risk difference scale cannot in general be estimated from case-control data. Rothman [6] proposed instead a measure of additive

interaction that used risk ratios rather than absolute risks. Let $RR_{x_1x_2} = \frac{p_{x_1x_2}}{p_{00}}$ be the risk ratio comparing the exposure category $X_1 = x_1, X_2 = x_2$ to $X_1 = 0, X_2 = 0$. Rothman [6] defined the relative excess risk due to interaction ($RERI$) as $RERI = RR_{11} - RR_{10} - RR_{01} + 1$. This $RERI$ measure is equal to the additive interaction for absolute risks, $p_{11} - p_{10} - p_{01} + p_{00}$, divided by p_{00} . The $RERI$ measure will be greater than zero if and only if the additive interaction is greater than zero; it will be equal to zero if and only if the additive interaction is equal to zero; it will be less than zero if and only if the additive interaction is less than zero. The $RERI$ measure thus allows one to assess additive interaction using risk ratios. If the outcome D is rare, or if incidence density sampling is used, odds ratios will approximate risk ratios and the $RERI$ measure can be assessed with case-control data. A number of estimation techniques have been developed for obtaining estimates and confidence intervals for $RERI$ [2, 7–9]. The $RERI$ measure is also useful for assessing evidence for mechanistic interaction within the sufficient cause and potential outcomes frameworks [3–5]. If $RERI > 0$

then there is evidence of “mechanistic” or “sufficient cause” interaction (i.e. there are individuals for whom the outcome occurs if both exposures are present but not if just one or the other is present) provided it can be assumed that neither exposure is ever preventive (only causative or neutral) for all individuals in the population, an assumption sometimes referred to as monotonicity. If $RERI > 1$ then there is evidence of such “mechanistic” or “sufficient cause” interaction even without such monotonicity assumptions.

Rothman [6] also proposed a measure which he called the attributable proportion defined as

$$AP = \frac{RERI}{RR_{11}} = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11}}.$$

The measure was intended to capture the proportion of the *disease* in the doubly exposed group that is due to the interaction between the two exposures. In his book [6], Rothman also defined an alternative measure also referred to there as the “attributable proportion” defined as:

$$AP^* = \frac{RERI}{RR_{11} - 1}.$$

This measure was intended to capture the proportion of the *effect* on the additive scale of both exposures combined that is due to the interaction between the two exposures. Although Rothman presented these measures using $RERI$ and risk ratios, they are applicable also to

cohort studies using risks; the former measure can be written as $AP = \frac{p_{11} - p_{10} - p_{01} + p_{00}}{p_{11}}$ and the latter as $AP^* = \frac{p_{11} - p_{10} - p_{01} + p_{00}}{p_{11} - p_{00}}.$

In contrasting the two measures, the former measure, AP , captures the proportion of the *disease* in the doubly exposed group that is due to the interaction; the latter measure, AP^* , captures the proportion of the *effect* of both exposures that is due to interaction. The former measure uses RR_{11} as the denominator; the latter measure uses $RR_{11} - 1$. The focus of the subsequent literature and reporting has been on the former measure. This was perhaps because greater emphasis was placed on this measure in Rothman's book and also because methods for statistical inference and software to estimate the attributable proportion and its confidence interval focused on the former measure [7, 8]. Of course, an investigator should select between the two based on substantive considerations, i.e. whether the proportion of disease in the doubly exposed group due to interaction is in view, or the proportion of the effect is in view. In most instances, however, both measures will likely be of interest and both could be reported. But often in practice, only the former measure is selected, essentially by default, and these considerations of interpretation of the two alternative measures are, for the most part, ignored. Here we will consider the properties of these two different attributable proportion measures. We will see that there are a number of attractive properties of using the latter measure that are not shared by former.

First, consider the setting in which the entirety of the effect were due to interaction. For example, suppose $RR_{11} = 3$ but $RR_{10} = RR_{01} = 1$ so that there is no effect of either of the exposures in the absence of the other. The effect of the exposures is only manifest when both exposures are present. When both exposures are present, the entirety of the effect is due to interaction. A researcher who does not distinguish between the attributable proportion

“for *risk* in the doubly exposed group” versus the attributable proportion for the “*effects* of both exposures” might think that the attributable proportion measure in this case should be 100 %. This might seem like it should be the natural reference for the maximum attributable proportion as, in the case considered here, there is no effect of the exposures except for the interaction. Using Rothman's primary definition we get an attributable proportion measure of:

$$AP = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11}} = \frac{3 - 1 - 1 + 1}{3} = \frac{2}{3} = 66.7\%.$$

We do not get 100 %. This is because we are using the denominator, RR_{11} , rather than, $RR_{11} - 1$. The focus is here on the proportion of the disease in the doubly exposed group due to interaction, not the proportion of the effect. Using the alternative attributable proportion definition we obtain

$$AP^* = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11} - 1} = \frac{3 - 1 - 1 + 1}{3 - 1} = \frac{2}{2} = 100\%$$

as might be expected. Again for this alternative definition, we are capturing the proportion of the *effect* due to interaction, not simply the proportion of the *disease* in the doubly exposed group due to interaction. If Rothman's primary definition, AP , is to be used then it must be kept in mind that the measure will not be 100 % even if the effects of both exposures are due entirely to interaction.

Next, consider what happens upon reversing the coding of the outcome. It is shown in Appendix 1, that with the alternative definition of the attributable proportion,

$AP^* = \frac{RERI}{RR_{11} - 1}$, reversing the coding of the outcome still gives the same attributable proportion measure as one obtains with the original coding. This might be of relevance, for example, if the outcome is survival versus death, and, in this setting, one might think that the attributable proportion should be the same irrespective of which of the two codings is chosen as the outcome. With Rothman's primary definition of the attributable proportion,

$AP = \frac{RERI}{RR_{11}}$, if we reverse the coding of the outcome, the proportion attributable to interaction changes. For example, suppose that the outcome D were death and the underlying outcome probabilities were $p_{00} = 0.01$, $p_{10} = 0.02$, $p_{01} = 0.02$, $p_{11} = 0.06$ with relative risks $RR_{10} = 2$, $RR_{01} = 2$, $RR_{11} = 6$. If we were to reverse the outcome to survival we would have corresponding outcome probabilities $p_{00} = 0.99$, $p_{10} = 0.98$, $p_{01} = 0.98$, $p_{11} =$

0.94 and relative risks of $RR_{10} = \frac{98}{99}$, $RR_{01} = \frac{98}{99}$, $RR_{11} = \frac{94}{99}$. For the alternative attributable

proportion definition, $AP^* = \frac{RERI}{RR_{11} - 1}$, using death as the outcome we would have an attributable proportion measure of

$$AP^* = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11} - 1} = \frac{6 - 2 - 2 + 1}{6 - 1} = \frac{3}{5} = 60\%, \text{ and using survival as the}$$

outcome the measure is still $\frac{94/99 - 98/99 - 98/99 + 1}{94/99 - 1} = \frac{-3/99}{-5/99} = 60\%$. For Rothman's primary attributable proportion definition, using death as the outcome we would have an attributable proportion measure of

$$AP = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11}} = \frac{6 - 2 - 2 + 1}{6} = \frac{3}{6} = 50\%$$
 and using survival as the outcome the measure would then become

$$\frac{94/99 - 98/99 - 98/99 + 1}{94/99} = \frac{-3/99}{94/99} = -3.19\%$$
 It may be thought desirable to attribute the same proportion of the effects of the exposures to interaction irrespective of whether death or survival is considered as the outcome.

Of course, in many settings there is a natural coding of the outcome (e.g. heart disease present versus absent, or cancer present versus absent) and we might not ever think to, or be interested in, recoding. Moreover, in a case-control context we cannot reverse the definition of the outcome and retain a rare outcome assumption, but the recoding of the outcome may sometimes be of interest in cohort studies. Specifically, the invariance of the alternative attributable proportion definition to the coding of the outcome has important implications for preventive exposures. Building on Rothman [10], Knol et al. [11] suggested that the interpretation of measures like *RERI* were clearest when the effects of each of the exposures in the absence of the other were positive and proposed a recoding strategy of exposures to ensure that that was so. In some settings one could alternatively recode the outcome to ensure this. Importantly, because of the invariance of the alternative attributable proportion definition, AP^* , to the coding of the outcome, the attributable proportion measure can potentially be used and easily interpreted for preventive exposures as well. Suppose our

relative risks were $RR_{10} = \frac{3}{4}$, $RR_{01} = \frac{3}{4}$, $RR_{11} = \frac{1}{4}$. The *RERI* measure is then

$$RR_{11} - RR_{10} - RR_{01} + 1 = \frac{1}{4} - \frac{3}{4} - \frac{3}{4} + 1 = -\frac{1}{4}$$
 and the alternative attributable proportion measure is

$$AP^* = \frac{1/4 - 3/4 - 3/4 + 1}{1/4 - 1} = \frac{-1/4}{-3/4} = \frac{1}{3} = 33.3\%$$
 essentially, each of the exposures considered alone reduces the relative risk by 1/4, from 1 to 3/4; if the effects were additive we would expect both exposures together to reduce the relative risk from 1 to 1/2; in fact, both exposures together reduce the relative risk by 3/4 from 1 to 1/4. Thus of this 3/4 point reduction in relative risk for both exposures together, a 1/2 point reduction is from combining each of the exposures considered alone and 1/4 is due to the interaction. The

proportion attributable to interaction is thus $\frac{1/4}{3/4} = 33.3\%$. This is what the alternative attributable proportion measure $AP^* = \frac{RERI}{RR_{11} - 1}$ indeed gives us. Moreover, if we were to recode the outcome so that the effects of the exposures were causative rather than preventive we would still obtain the same attributable proportion measure of $AP^* = 33.3\%$.

Third, Skrondal [12] criticized Rothman's original attributable proportion measure (and also *RERI*) because, in the presence of covariates, if the risks follow a linear risk model that is additive in the covariates, $P(D=1 | X_1 = x_1, X_2 = x_2, C = c) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + \alpha_4 c$, then, although the additive interaction, $p_{11} - p_{10} - p_{01} + p_{00} = \alpha_3$, does not vary across strata of the covariates, Rothman's primary attributable proportion measure,

$$AP = \frac{RERI}{RR_{11} - 1} = \frac{\alpha_3}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 c},$$
 does vary across strata of the covariates. One may or may not think that this is an important criticism of the attributable proportion measure AP ; however, it is interesting to note that the alternative attributable proportion measure,

$$AP^* = \frac{RERI}{RR_{11} - 1} = \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3},$$
 does not vary with the covariates and thus circumvents this criticism entirely.

We have seen then a number of attractive properties of taking the denominator of attributable proportion for additive interaction as $RR_{11} - 1$ rather than RR_{11} so that the

alternative attributable proportion $AP^* = \frac{RERI}{RR_{11} - 1}$ is used, rather than $AP = \frac{RERI}{RR_{11}}$. When all of the effect is due to interaction the alternative attributable proportion is then 100 % as might be expected and not some number less than 100 %; the alternative measure is also invariant to the coding of the outcome and can be used for preventive exposures; and finally, the alternative measure, unlike the Rothman's primary measure, will not vary with covariates under a linear risk model. All of these properties follow from focusing on the proportion of the *effect* of both exposures on the additive scale that is attributable to interaction, rather than the proportion of *disease* in the doubly exposed group due to interaction. Once again, in choosing between, reporting, and interpreting these measures, one must consider whether the proportion of *disease* in the doubly exposed group due to interaction, or the proportion of the *effect* of both exposures that is attributable to interaction is of interest. Ideally both could perhaps be reported. However, researchers should realize that there is in fact a distinction, and the measures need to be interpreted accordingly. As seen here, the alternative attributable proportion measure has a number of attractive properties. In any case, when using attributable proportion measures for interaction in practice, further refinement in terminology would perhaps make both the choice, and the interpretation, of

the measures more transparent. The measure $AP = \frac{RERI}{RR_{11}}$ could potentially be referred to as the “attributable proportion of (doubly exposed) risk” and the measure $\text{measure} \frac{RERI}{RR_{11} - 1}$ as the “attributable proportion of (joint) effects.”

To facilitate use of this alternative attributable proportion definition, $AP^* = \frac{RERI}{RR_{11} - 1}$, standard errors for this measure, using logistic regression assuming a rare outcome are derived in the Online Appendix, and SAS code to obtain estimates and confidence intervals are provided in Appendix 2 below.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix 1

Invariance to recoding the outcome

Here we show that the alternative definition of the attributable proportion, $AP^* = \frac{RERI}{RR_{11} - 1}$, is invariant to the recoding of the outcome, whereas Rothman's primary definition,

$AP = \frac{RERI}{RR_{11}}$, is not. The definition of the alternative attributable proportion is [6]:

$$AP^* = \frac{RERI}{RR_{11} - 1} = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11} - 1} = \frac{\frac{p_{11}}{p_{00}} - \frac{p_{10}}{p_{00}} - \frac{p_{01}}{p_{00}} + 1}{\frac{p_{11}}{p_{00}} - 1} = \frac{(p_{11} - p_{10} - p_{01} + p_{00})}{p_{11} - p_{00}}.$$

If we reverse the coding of the outcome, the relative risks for the categories $(X_1 = 1, X_2 = 1)$, $(X_1 = 1, X_2 = 0)$, $(X_1 = 0, X_2 = 1)$ become respectively $\frac{1 - p_{11}}{1 - p_{00}}$, $\frac{1 - p_{10}}{1 - p_{00}}$, and $\frac{1 - p_{01}}{1 - p_{00}}$, and the alternative attributable proportion measure is then

$$AP^* = \frac{\frac{1-p_{11}}{1-p_{00}} - \frac{1-p_{10}}{1-p_{00}} - \frac{1-p_{01}}{1-p_{00}} + 1}{\frac{1-p_{11}}{1-p_{00}} - 1} = \frac{(1 - p_{11}) - (1 - p_{10}) - (1 - p_{01}) + (1 - p_{00})}{(1 - p_{11}) - (1 - p_{00})} = \frac{(p_{11} - p_{10} - p_{01} + p_{00})}{p_{11} - p_{00}}$$

which is the same measure we obtained under the original coding of the outcome.

For Rothman's primary definition of the attributable proportion, $AP = \frac{RERI}{RR_{11}}$, under the original coding we have that this is

$$AP = \frac{RERI}{RR_{11}} = \frac{RR_{11} - RR_{10} - RR_{01} + 1}{RR_{11}} = \frac{\frac{p_{11}}{p_{00}} - \frac{p_{10}}{p_{00}} - \frac{p_{01}}{p_{00}} + 1}{\frac{p_{11}}{p_{00}}} = \frac{(p_{11} - p_{10} - p_{01} + p_{00})}{p_{11}}.$$

If we reverse the coding of the outcome, the relative risks for the categories $(X_1 = 1, X_2 = 1)$, $(X_1 = 1, X_2 = 0)$, $(X_1 = 0, X_2 = 1)$ become respectively $\frac{1 - p_{11}}{1 - p_{00}}$, $\frac{1 - p_{10}}{1 - p_{00}}$, and $\frac{1 - p_{01}}{1 - p_{00}}$, and Rothman's primary attributable proportion measure is then

$$AP = \frac{\frac{1-p_{11}}{1-p_{00}} - \frac{1-p_{10}}{1-p_{00}} - \frac{1-p_{01}}{1-p_{00}} + 1}{\frac{1-p_{11}}{1-p_{00}}} = \frac{(1 - p_{11}) - (1 - p_{10}) - (1 - p_{01}) + (1 - p_{00})}{(1 - p_{11})} = \frac{-(p_{11} - p_{10} - p_{01} + p_{00})}{1 - p_{11}}$$

so that the measure reverses sign and is of a very different magnitude.

Appendix 2

SAS and Stata code to implement the alternative attributable proportion measure

Suppose the outcome is in variable 'd', the first exposure in variable 'g' and the second exposure in variable 'e' with three covariates 'c1 c2 c3'. The following SAS code will

estimate the alternative attributable proportion measure, $AP^* = \frac{RERI}{RR_{11} - 1}$, and its confidence interval using the delta method (see Online Appendix):

```
proc logistic descending data=mydata outest=myoutput covout;
  model d=g e g*e c1 c2 c3;
run;
data PAoutput;
  Set myoutput;
  array mm {*} _numeric_;
```

```

b0=lag4(mm[1]);
b1=lag4(mm[2]);
b2=lag4(mm[3]);
b3=lag4(mm[4]);
v11=lag2(mm[2]);
v12=lag(mm[2]);
v13=mm[2];
v22=lag(mm[3]);
v23=mm[3];
v33=mm[4];
f1=(exp(b1)+(exp(b2)-2)*exp(b1+b2+b3))/((exp(b1+b2+b3)-1)*(exp(b1+b2+b3)-1));
f2=(exp(b2)+(exp(b1)-2)*exp(b1+b2+b3))/((exp(b1+b2+b3)-1)*(exp(b1+b2+b3)-1));
f3=((exp(b1)+exp(b2)-2)*exp(b1+b2+b3))/((exp(b1+b2+b3)-1)*(exp(b1+b2+b3)-1));
vINT=v11*f1*f1 + v22*f2*f2 + v33*f3*f3 + 2*v12*f1*f2 + 2*v13*f1*f3 + 2*v23*f2*f3;
PaINT=(exp (b2+b1+b3)-exp(b1)-exp(b2)+1)/((exp (b1+b2+b3)-1);
se_PaINT=sqrt(vINT);
ci95_lINT=PaINT-1.96*se_PaINT;
ci95_uINT=PaINT+1.96*se_PaINT;
keep PaINT ci95_lINT ci95_uINT;
if _n_=5;
run;
proc print data=PAoutput;
    Var PaINT ci95_lINT ci95_uINT;
run;

```

The following Stata code will estimate the alternative attributable proportion measure and its confidence interval:

```

generate Ige = g*e
logit d g e Ige c1 c2 c3
nlcom (exp (_b [g] +_b [e] +_b [Ige])-exp (_b [g])-exp (_b [e]) +1)/(exp (_b [g] +_b [e] +_b [Ige])-1)

```

References

1. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol.* 1980; 112(4): 467–470. [PubMed: 7424895]
2. Rothman, KJ.; Greenland, S.; Lash, TL. *Modern epidemiology.* 3rd. Philadelphia: Lippincott Williams & Wilkins; 2008.
3. VanderWeele TJ, Knol MJ. A tutorial on interaction *Epidemiol Methods.* in press.
4. Rothman KJ. Causes. *Am J Epidemiol.* 1976; 104:587–592. [PubMed: 998606]
5. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology.* 2009; 20(1):6–13. [PubMed: 19234396]
6. Rothman, KJ. *Modern epidemiology.* 1st. Boston: Little, Brown and Company; 1986.
7. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology.* 1992; 3(5):452–456. [PubMed: 1391139]
8. Lundberg M, Fredlund P, Hallqvist J, Diderichsen F. A SAS program calculating three measures of interaction with confidence intervals.
9. VanderWeele TJ, Vansteelandt S. A weighting approach to causal effects and additive interaction in case–control studies: marginal structural linear odds models. *Am J Epidemiol.* 2011; 174:1197–1203. [PubMed: 22058231]
10. Rothman, KJ. *Epidemiology: an introduction.* Oxford; Oxford University Press; 2002.

11. Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. *Eur J Epidemiol.* 2011; 26:433–438. [PubMed: 21344323]
12. Skrdonal A. Interaction as departure from additivity in case–control studies: a cautionary note. *Am J Epidemiol.* 2003; 158(3):251–258. [PubMed: 12882947]