Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Assuming No Modification

Let us consider the study of a disease (levels: 0 and 1) – exposure (levels: 0 and 1) relationship with a single potential modifier/confounder that we will call the 'strata' (levels: 0 and 1). The stratum specific probabilities of disease will be called: p_{ij} for the ith exposure and the jth stratum. For this illustration, we will suppose that these population characteristics are not troubled by further confounding or modification.

Rate differences:

If there is no RD modification, then $p_{10}-p_{00}=p_{11}-p_{01}$ which is then the rate difference RD. This is the 'correct' RD.

Now we are obliged to consider 'crude' determinations. We wish to consider 2 weighted sums:

$w_0 p_{01}+(1-w_0) p_{00}$ and $w_1 p_{11}+(1-w_1) p_{10}$

where, for the study under consideration, w_0 is the proportion among the unexposed of subjects in strata 1 and w_1 is the proportion among the exposed of subjects in strata 1. We will choose to think of these proportions as specific to the study at hand and not necessarily related to any population.

The challenge is to determine whether these 2 weighted sums distort the message that is available from the stratum specific analogues.

We can note that:

$p_{10}=RD+p_{00}$ and $p_{11}=RD+p_{01}$ since there is no RD modification.

Accordingly we wish to consider:

$w_0 p_{01}+(1-w_0) p_{00}$ and $w_1(RD+p_{01})+(1-w_1)(RD+p_{00})$

Taking the second minus the first gives us [say cRD] :

$$RD+(w_1-w_0)(p_{01}-p_{00})$$

Notice that the stratum specific difference RD differs from the comparison of the 2 weighted sums by:

$$(w_1-w_0)(p_{01}-p_{00})$$

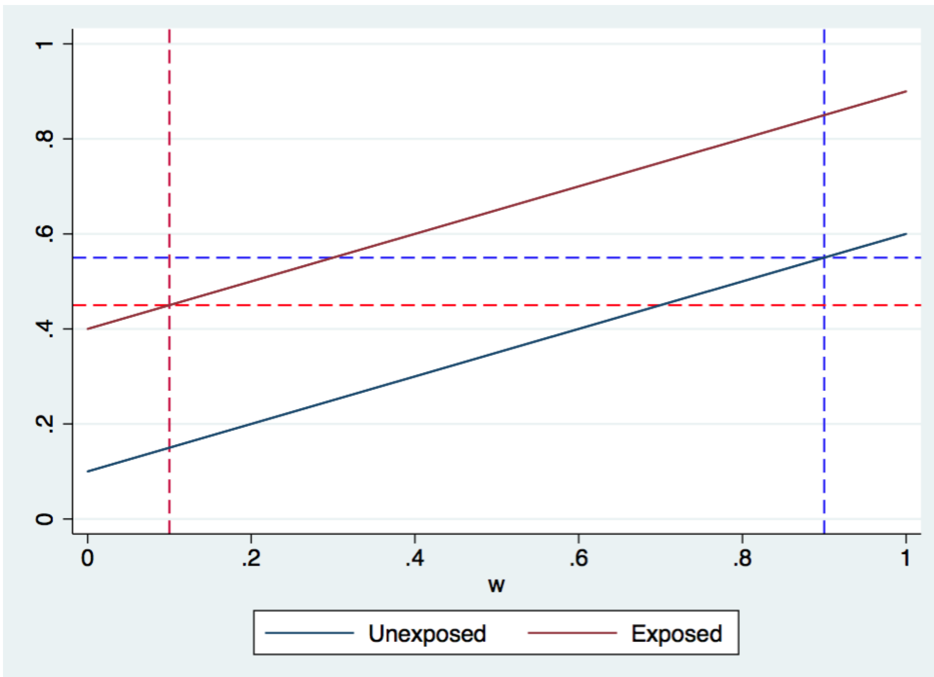
So both the terms (w_1-w_0) and $(p_{01}-p_{00})$ must be nonzero, to have a nonzero difference.

This observation is the same as the common statement that:

1) the weights must be different in the 2 exposure groups

and
2) the 'risk' in the absence of exposure must be different in the 2 strata.
Notice that since there is no RD modification, this statement is the same for those exposed.
$(p_{11}-p_{10})=(p_{01}-p_{00})$

RD Example 1
Using: $p_{00}=0.1\ p_{01}=0.6\ p_{10}=0.4\ p_{11}=0.9$ so that RD=0.3
Now consider $w_0=0.9\ w_1=0.1$
Then $RD+(w_1-w_0)(p_{01}-p_{00})=0.3+(0.1-0.9)*(0.6-0.1)=0.3-0.8*0.5=-0.1$
The visual below illustrates this matter.



Naming $p_i(w)=p_{i1}w+p_{i0}(1-w)$
So, without modification we have $p_1(w)=RD+p_0(w)$

To see a sign change in cRD, the weights satisfy $w_1-w_0<-\dfrac{RD}{p_{01}-p_{00}}$

In the example, $w_1-w_0<-0.3/0.5=-0.6$ where w_1 could be as small as zero and w_0 could be as large as one.

To see cRD = k* RD we have $w_1-w_0=(k-1)\dfrac{RD}{p_{01}-p_{00}}$

Rate Ratios:
If there is no RR modification, then $p_{10}/p_{00}=p_{11}/p_{01}$ which is then the rate ratio RR.
We can note that:

$p_{10}=RR*p_{00}$ and $p_{11}=RR*p_{01}$ since there is no modification.

Accordingly we wish to consider:

$w_0 p_{01} + (1-w_0) p_{00}$ and $w_1(RR * p_{01}) + (1-w_1)(RR * p_{00})$

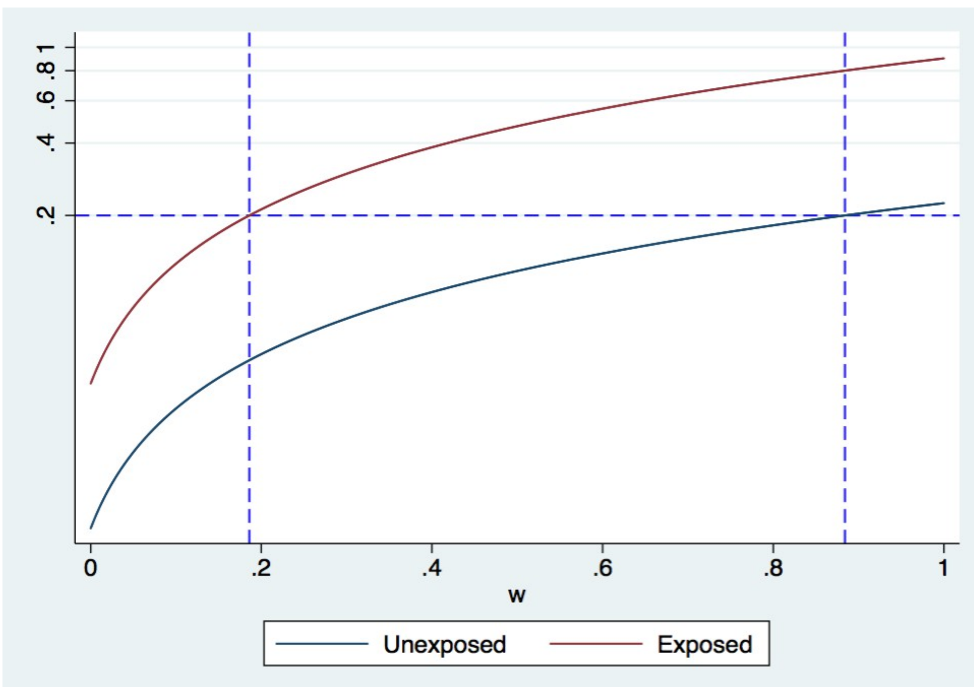
Taking the second divided by the first gives us [say cRR]:

$RR * \dfrac{w_1 p_{01} + (1-w_1) p_{00}}{w_0 p_{01} + (1-w_0) p_{00}}$ which equals RR when $(w_1 - w_0)(p_{01} - p_{00}) = 0.$

This is the same condition as for the RD.
Since $(p_{11} - p_{10}) = RR * (p_{01} - p_{00})$ we have a similar statement as with RD.

Now we have $\log p_1(w) = \log RR + \log p_0(w)$

RR Example 1
Using: $p_{00} = 0.01$ $p_{01} = 0.225$ $p_{10} = 0.04$ $p_{11} = 0.9$ so that RR=4
The graph shows a case when cRR=1.



Odds Ratios:

If there is no OR modification, then $\dfrac{p_{10}/(1-p_{10})}{p_{00}/(1-p_{00})} = \dfrac{p_{11}/(1-p_{11})}{p_{01}/(1-p_{01})}$ which is the odds ratio OR.

$p_{10}/(1-p_{10}) = \mathrm{OR} * p_{00}/(1-p_{00})$

$p_{11}/(1-p_{11}) = \mathrm{OR} * p_{01}/(1-p_{01})$

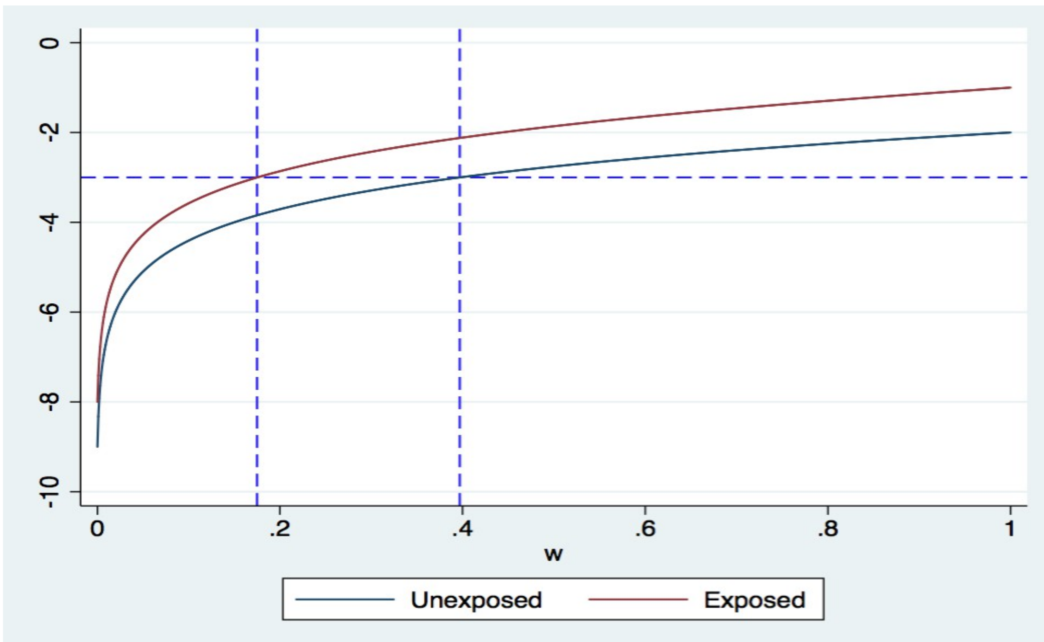
But we do not have $\log \dfrac{p_1(w)}{1-p_1(w)} = \log \mathrm{OR} + \log \dfrac{p_0(w)}{1-p_0(w)}$

In fact, $\log\frac{p_1(w)}{1-p_1(w)}$ - $\log\frac{p_0(w)}{1-p_0(w)}$ is a function of w.

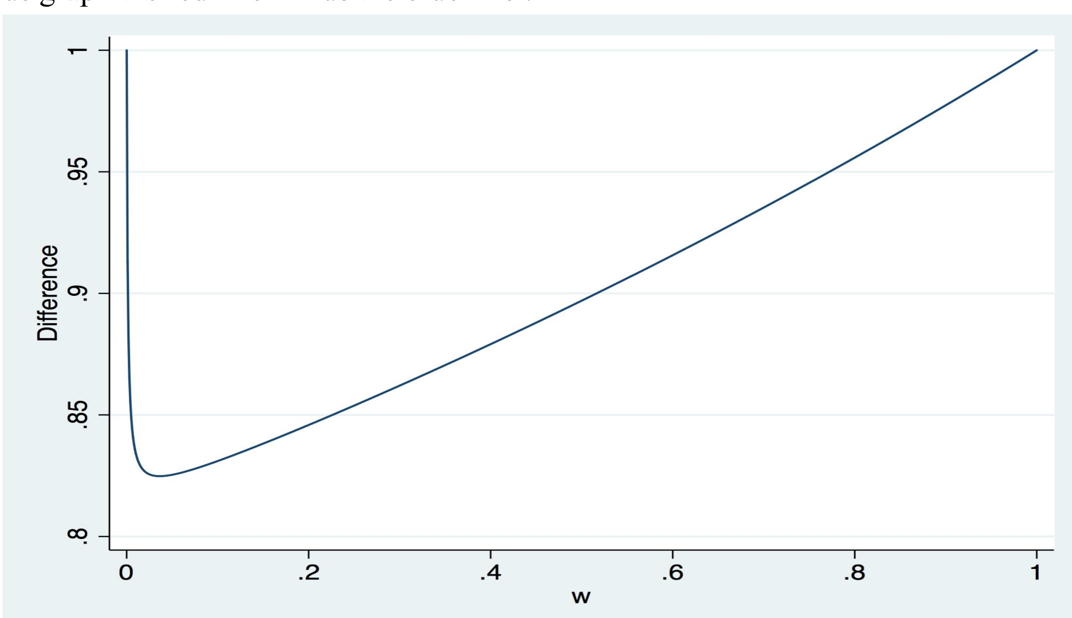
OR Example 1

Using: $\log\frac{p_{00}}{1-p_{00}}=-9$ $\log\frac{p_{10}}{1-p_{10}}=-8$ $\log\frac{p_{01}}{1-p_{01}}=-2$ $\log\frac{p_{11}}{1-p_{11}}=-1$ so that log OR =1

So there is no OR modification but but there is modification with RD, RR and HR. $RD_0=.000212$ $RD_1=.149738$ $RR_0=2.717706$ $RR_1=2.256165$ $HR_0=.999788$ $HR_1=.829997$



Now let us graph the red line minus the blue line :



So $\log\frac{p_1(w)}{1-p_1(w)}$ - $\log\frac{p_0(w)}{1-p_0(w)}$ is not equal to the assumed common log OR of 1 but can less than 0.85 in this example. Notice, then, that this phenomenon is not confounding. It goes by many names.

Maybe calling it attenuation would be the clearest name for it. It can be shown that the difference $\log\frac{p_1(w)}{1-p_1(w)}$ - $\log\frac{p_0(w)}{1-p_0(w)}$ will always be less than the assumed common log OR.

Other names seen in the literature are 'non-collapsibility' and even 'non-linearity'. Sometimes this non-constant difference is illustrated with examples where modification is also present. This makes the issue harder to parse.

The first publication on this matter appears to be :
Myra L. Samuels : Biometrika, Vol. 68, No. 3 (Dec., 1981), pp. 577-588.

Myra Samuels (1940-1992) was a member of the Department of Veterinary Pathobiology at Purdue University. She completed her PhD (Statistics) University of California, Berkeley with supervisor Jerzy Neyman.

In her paper, Samuels 'applied' Jensen's inequality. Very cool !
Many authors have written about this matter since Samuels.

The study of the magnitude of the attenuation is an active area of research.
So when will the attenuation matter?
If the HR is "near" one, then attenuation will be "minor".

Let $p_{ij}=x\,q_{ij}$ For given q_{ij}, the RR will be constant As $x\downarrow 0$, $HR\rightarrow 1$ and so OR $\rightarrow$ RR

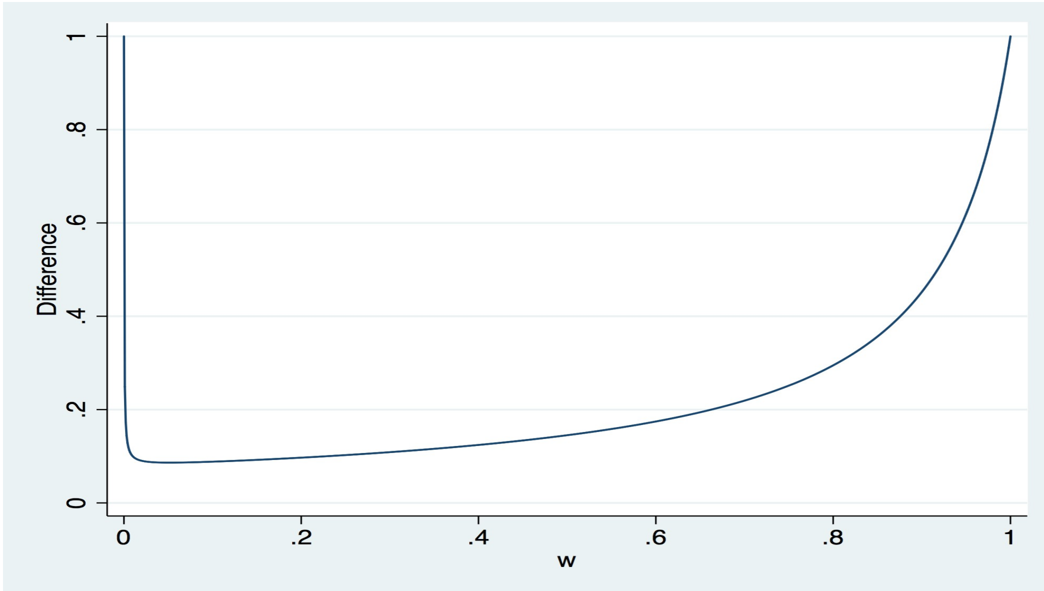
Recently, Diego Nobrega and I have developed a formula for the maximum attenuation.

With $a=p_{00}$ $b=p_{10}$ $c=p_{01}$ $d=p_{11}$
$e=bc-ad$ $f=(1-b)(1-c)-(1-a)(1-d)$
$g=(c-a)(d-b)$ $h=ef/g$
$r=(d-c)-(b-a)$
we get that :

$$A_{max}=\log\left(\left((1-w)a+wc\right)^{-1}-1\right)-\log\left(\left((1-w)b+wd\right)^{-1}-1\right)$$

when $w = \frac{\sqrt{h}-(b-a)}{r}$

Here is another even more extreme example :



All of the development in this session is for the population domain.

With actual data,

1) We get estimates of the probabilities.
2) We get the “actual” fractions.
3) There are finite sets of possibilities for the estimates and the actual fractions.
4) There are the practical issues of modification and confounding.
5) There are the merits and demerits for exploring counterfactuals.