

# Effect size measures in genetic association studies and age-conditional risk prediction

Hon-Cheong So and Pak C. Sham

## Supplementary materials

Please note that relevant R programs are available at

<http://sites.google.com/site/honcheongso/software/gene-pred>

### Detailed mathematical derivations for prevalence odds ratio

The derivations below are adapted from Alho [1]. The probability that a person survives up to age  $a$  without the disease of interest,  $S(a)$ , can be expressed as

$$S(a) = \exp\left(-\int_0^a (\mu(x) + \lambda(x)) dx\right)$$

As pointed out by one reviewer, for functions of the form  $\eta(a) = \exp\int_0^a g(x)dx$ , the

derivative  $\frac{d\eta(a)}{da} = \eta(a)g(a)$ , which means that interestingly the rate of change of  $\eta(a)$  is

proportional to  $\eta(a)$  itself. In the above formula,  $\frac{dS(a)}{da} = -S(a)[\mu(a) + \lambda(a)]$ , therefore

the rate of decrease in  $S(a)$  with respect to the age is proportional to  $S(a)$  (and the hazard function  $[\mu(a) + \lambda(a)]$ ).

Suppose that people are born at a rate of  $Be^{\rho t}$ . At exact time  $t$ , the expected size of the population who are free of the disease of interest is

$$\begin{aligned} H(t) &= \int_0^\infty Be^{\rho(t-a)} S(a) da \\ &= Be^{\rho t} \int_0^\infty \exp\left(-\int_0^a (\rho + \mu(x) + \lambda(x)) dx\right) da \end{aligned}$$

The age distribution of the disease-free individuals is given by

$$f_\rho(a) = \frac{Be^{\rho(t-a)} S(a)}{H(t)}$$

If the population is assumed to be stationary, then  $\rho=0$  and the above expressions may be simplified to

$$H(t) = B \int_0^\infty \exp\left(-\int_0^a (\mu(x) + \lambda(x)) dx\right) da$$

and

$$\begin{aligned} f_0(a) &= \frac{BS(a)}{H(t)} \\ &= \frac{\exp\left(-\int_0^a (\mu(x) + \lambda(x)) dx\right)}{\int_0^\infty \exp\left(-\int_0^a (\mu(x) + \lambda(x)) dx\right) da} \end{aligned}$$

Also also derived that the expected size of the population with disease at an exact time  $t$  is

$$I(t) = Be^{\rho t} \int_0^\infty \exp\left(-\int_0^z (\rho + \mu(x) + \lambda(x)) dx\right) \lambda(z) D_\rho(z) dz$$

where  $D_\rho(z)$  is the discounted expected duration of disease given by

$$D_\rho(z) = \int_0^\infty \exp\left(-\int_z^a (\rho + \mu(x) + \alpha(x-z, z)) dx\right) da$$

The discount rate depends on  $z$ .

### Disease odds ratio and exposure odds ratio

It is well-known that the disease odds ratio is equal to the exposure odds ratio. Note that in case-control studies we can only obtain the exposure odds ratio because the subjects are sampled based on whether they are affected or not. The derivation is based on simple algebra and may be found in many texts [e.g. [2] ]

	Disease(D)	No disease( $\bar{D}$ )	Total
Exposed (e)	a	b	p
Not exposed( $\bar{e}$ )	c	d	q
Total	r	s	T

The proof is summarized below using notations in the above table:

$$\text{Disease odds ratio} = \frac{\text{odds}(D|e)}{\text{odds}(D|\bar{e})} = \frac{\Pr(D|e) / \Pr(\bar{D}|e)}{\Pr(D|\bar{e}) / \Pr(\bar{D}|\bar{e})} = \frac{(a/p) / (b/p)}{(c/q) / (d/q)} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

$$\text{Exposure odds ratio} = \frac{\text{odds}(e|D)}{\text{odds}(e|\bar{D})} = \frac{\Pr(e|D) / \Pr(\bar{e}|D)}{\Pr(e|\bar{D}) / \Pr(\bar{e}|\bar{D})} = \frac{(a/r) / (c/r)}{(b/s) / (d/s)} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Note that disease odds ratio  $\frac{odds(D|e)}{odds(D|\bar{e})} = \frac{I_e(t)/H_e(t)}{I_{ue}(t)/H_{ue}(t)}$

Using a similar argument, we can prove that if the controls are unselected (i.e. population controls are used), the resulting exposure odds is equal to the prevalence ratio of the exposed group to the unexposed group:

$$\frac{o(E|D)}{o(E)} = \frac{\Pr(D|E)}{\Pr(D|\bar{E})}$$

The proof is as follows:

$$\begin{aligned} \frac{o(E|D)}{o(E)} &= \frac{\Pr(E|D)/\Pr(\bar{E}|D)}{\Pr(E)/\Pr(\bar{E})} \\ &= \frac{\frac{\Pr(ED)}{\Pr(D)} / \frac{\Pr(\bar{E}D)}{\Pr(D)}}{\Pr(E)/\Pr(\bar{E})} \\ &= \frac{\Pr(ED)/\Pr(\bar{E}D)}{\Pr(E)/\Pr(\bar{E})} \\ &= \frac{\Pr(ED)}{\Pr(E)} / \frac{\Pr(\bar{E}D)}{\Pr(\bar{E})} \\ &= \frac{\Pr(D|E)}{\Pr(D|\bar{E})} \end{aligned}$$

This relation is used when we derive the genotype-specific risk relative to the general population.

### **Solution to OR compared to the general population**

For simplicity, we employ the following notations:

$$\begin{aligned} p &= \Pr(D|aa) / [1 - \Pr(D|aa)] \\ r &= \Pr(D|Aa) / [1 - \Pr(D|Aa)] \\ q &= \Pr(D|AA) / [1 - \Pr(D|AA)] \end{aligned}$$

Hence

$$\Pr(D|aa) = \frac{p}{1+p} \quad \Pr(D|Aa) = \frac{r}{1+r} \quad \Pr(D|AA) = \frac{q}{1+q}$$

We have

$$\Pr(D) = \frac{p}{1+p} \Pr(aa) + \frac{r}{1+r} \Pr(Aa) + \frac{q}{1+q} \Pr(AA)$$

Since

$$OR_2 = \frac{r}{p} \quad OR_3 = \frac{q}{p}$$

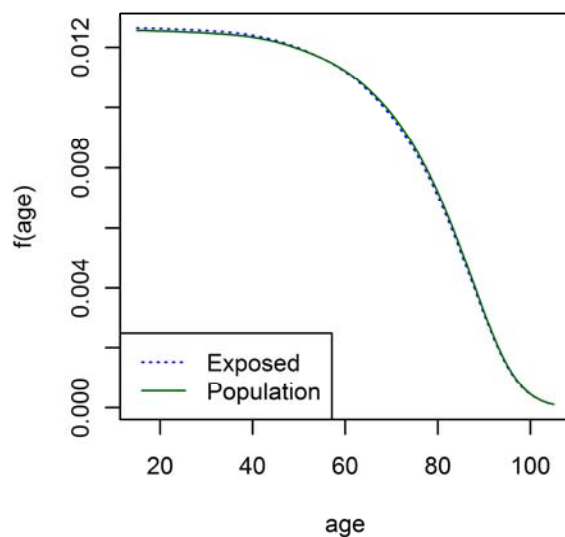
$\Pr(D)$  may be expressed as an equation in  $p$  only:

$$\Pr(D) = \frac{p}{1+p} \Pr(aa) + \frac{OR_2 p}{1+OR_2 p} \Pr(Aa) + \frac{OR_3 p}{1+OR_3 p} \Pr(AA)$$

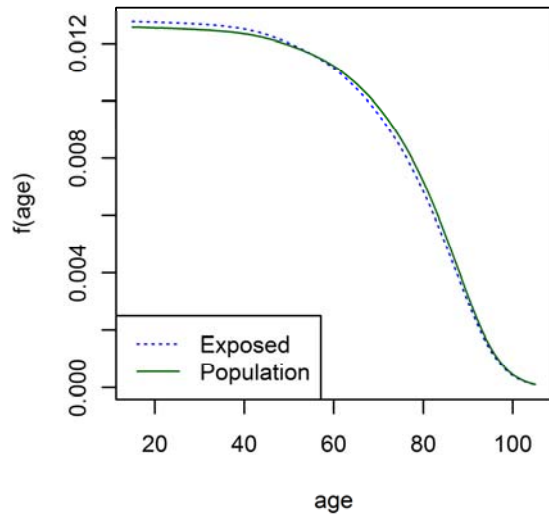
The equation may be solved numerically, for example by the uniroot function in R.

## Supplementary figures

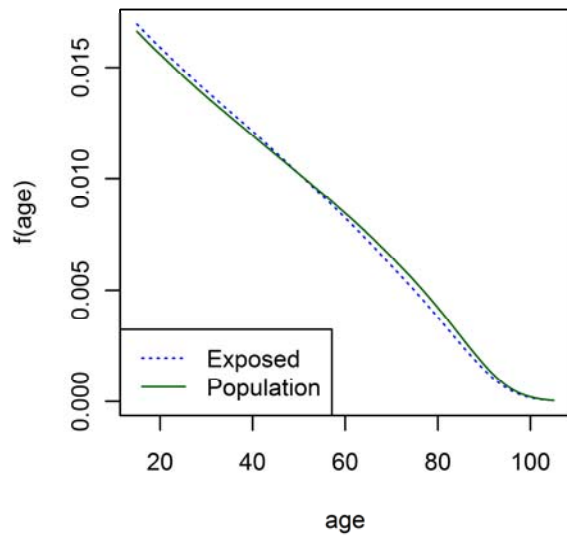
Supp Figure 1 Age distribution of disease-free individuals in the exposed group and in the general population for different intrinsic growth rates ( $\rho$ ) and hazard ratio(HR)



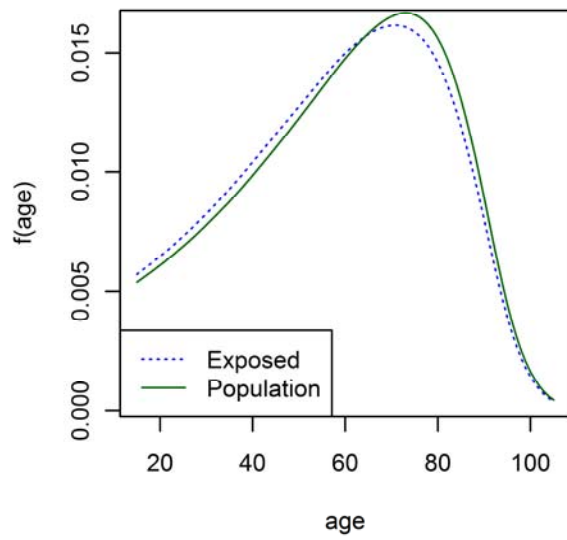
$\rho=0$  , HR= 1.2



$\rho=0$ , HR=1.5



$\rho = 0.0125$ , HR=2



$\rho = -0.0125$ ,  $HR=2$

References:

- 1 Alho JM: On prevalence, incidence, and duration in general stable-populations. *Biometrics* 1992;48:587-592.
- 2 Tsuang MT, Tohen M: Textbook in psychiatric epidemiology, ed 2nd. New York, Wiley-Liss, 2002.