**Appendix 4. Effect modification proof**

|  |  |  |  |
| --- | --- | --- | --- |
| **DAG** | **SCC model equivalent of DAG** | **Parameter inputs from simulation** | **True**  **effects** |
|  |  | Prevalence of exogenous variables in the simulation:  pZ1a=0.04  pZ2=0.1  pZ1b=0.1  pAPOE4=0.25  pLowEd=0.7 | Total sample:  RR=1.58  RD=0.058  APOE4+ stratum  RR=2.22  RD=0.122  APOE4- stratum  RR=1.36  RD=0.036 |

**I. Estimates in the total sample**

We first compute the crude risk ratio and risk difference in the total sample as follows:

)

The numerator and denominator quantities can be derived from the relevant causal components in Figure 6B and their parameter values specified in the simulation. For example, given an individual has low education, they will get dementia if they have either components Z1a, Z2, or APOE4&Z1b. Likewise, given an individual does not have low education, they will get dementia only if they have component Z2. The values of these component causes are specified above.

*Thus:*

= )

=

= 0.04+0.1+(0.25\*0.1)–(0.04\*0.1)–(0.04\*0.25\*0.1)–(0.1\*0.25\*0.1)+(0.04\*0.1\*0.25\*0.1)

= 0.1576

=

= 0.1

*And:*

These values for the crude RR and RD are unbiased (i.e., equal to the true effects specified in the simulation); the crude estimate is unbiased because of exchangeability of response types between exposed and unexposed, as shown in the main text Table 5.

**II. Estimates within strata of the effect modifier, APOE4**

To compute the crude risk ratio and risk difference among those with and without APOE4, we follow the same procedures as in Part I, only we condition on having and not having APOE4.

*Among those with APOE4:*

=

,

= 0.04 + 0.1 + 0.1 – (0.04\*0.1) – (0.04\*0.1) – (0.1\*0.1) + (0.04\*0.1\*0.1)

= 0.2224

, APOE4) =

and

= 0.1

*Among those without APOE4:*

=

= 0.04 + 0.1 - (0.04\*0.1)

= 0.136

, ) =

= 0.1

Again, the values for the RR and RD obtained within strata of APOE ε4 are unbiased (i.e., equal to the true effects specified in the simulation). As with the overall sample, there is exchangeability between exposed and unexposed within strata of APOE ε4, as shown in Table 5 in the main text. While possible to compute from the formulas above, Table 5 also more easily shows the proportions of doomed and causal types in each strata of exposure and APOE ε4, showing the higher prevalence of dementia in the exposed in the APOE ε4+ strata relative to the exposed in the APOE ε4- strata is due to a larger proportion of causal response types.