Supplementary information S3 | **Causal structure and diagrams**

Directed acyclic graphs (DAGs) are useful for depicting causal structure in epidemiologic settings. A DAG is composed of variables (nodes), both measured and unmeasured, and arrows (directed edges).

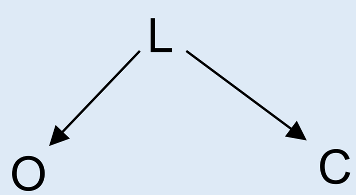
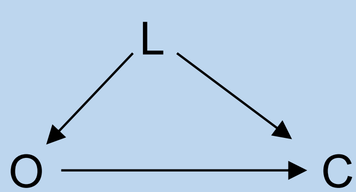
A causal DAG is one in which 1) the arrows can be interpreted as direct causal effects, and 2) all common causes of any pair of variables are included on the graph. Causal DAGs are acyclic because a variable cannot cause itself, either directly or through other variables.

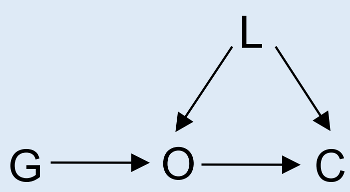
In the causal DAGs below, the figures presents the following dichotomous (for simple illustration) variables: O (being obese), L (having a westernised lifestyle of high energy intake and low physical activity); G (genetic determinant of obese state); and C (incident cancer).

Besides representing causal relations, causal DAGs also encode the causal determinants of statistical associations. For the purpose of the adiposity-cancer mechanistic discussion, we will consider two ways by which statistical associations can occur and under what condition a casual association can be inferred:[1]

First, a statistical association between O and C can be observed if there is a true causal effect from O on C.

Second, a statistical association between O and C can be observed if there exist common causes both of O and C.

Cause and effect  	In this hypothetical DAG, there is a 'true' causal effect between being obese and the occurrence of cancer
Common cause (no causal effect of obesity on cancer)  	In this hypothetical DAG, there is no causal effect from O on C (there is no arrow from O to C) but a statistical association will be observed between being obese and the occurrence of cancer. There is a common cause for O and C which is a westernized lifestyle (L) which produces the observed statistical association between O and C. In this scenario, adjusting for L by conditioning on L would remove the spurious association between O and C and reveal that there is no causal association.
Common cause (causal effect of obesity on cancer)  	In this hypothetical DAG, there is a causal association between being obese and the occurrence of cancer, but there is also a causal association between westernized lifestyle (L) and obesity and between L and cancer incidence. In this scenario, after appropriate adjustment of L, an association will remain between O and C which reflects the strength of the causal association. However, this will only hold true, if L was measured without measurement error. In the presence of measurement

	error, adjustment for L will not result in an unbiased estimate of the strength of the causal association of O on C.
Mendelian randomization 	In this hypothetical DAG, the genetic variant, G, takes the role of an instrumental variable (due to the random assignment of alleles in gamete formation). In this DAG the causal effect of G on O and of G on C can be estimated because they are not affected by confounding of L. Although the part of the DAG with only O, C and L looks similar to the DAG discussed above, it will be possible – even without measuring L - to retrieve an unbiased estimate of the strength of the causal effect of O on C by using the causal effects of G on O and of G on C if certain additional assumptions hold.[2]

References

1. Hernan, M.A., Hernandez-Diaz, S. & Robins, J.M. A structural approach to selection bias. *Epidemiology* **15**, 615-25 (2004).

2. Hernan, M.A. & Robins, J.M. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* **17**, 360-72 (2006).