Interactions and collider bias in gene-environment case-only data

Félix Balazard[1 2],S Le Fur[2], Isis-Diab collaborative group, P Bougnères[2], A-J Valleron[2]

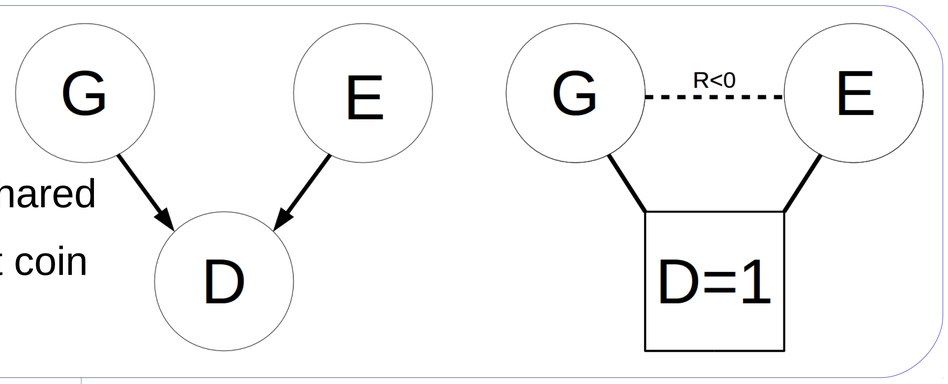
[1] : Laboratoire de Statistique Théorique et Appliquée, UPMC [2] : Unité 1169 INSERM

Interest of case-only data

Data are often easily available in patients of complex diseases. Selection of controls, a sensitive process, can be avoided.

Collider bias

It is the negative correlation that appears between causes when conditioning on a shared consequence [1], eg E and G independent coin toss and D=1 if at least one heads.



Two complementary methodologies

Case-only design for gene-environment interaction (CODGEI) [2]

Assumptions

- Independence between G and E in general population.
- *Rare disease assumption (implies no collider bias).*

Goal

Identify GxE interactions.

Disease as collider (DAC) NEW !!

Assumptions

- Independence between G and E in general population.
- No interactions.

Goal

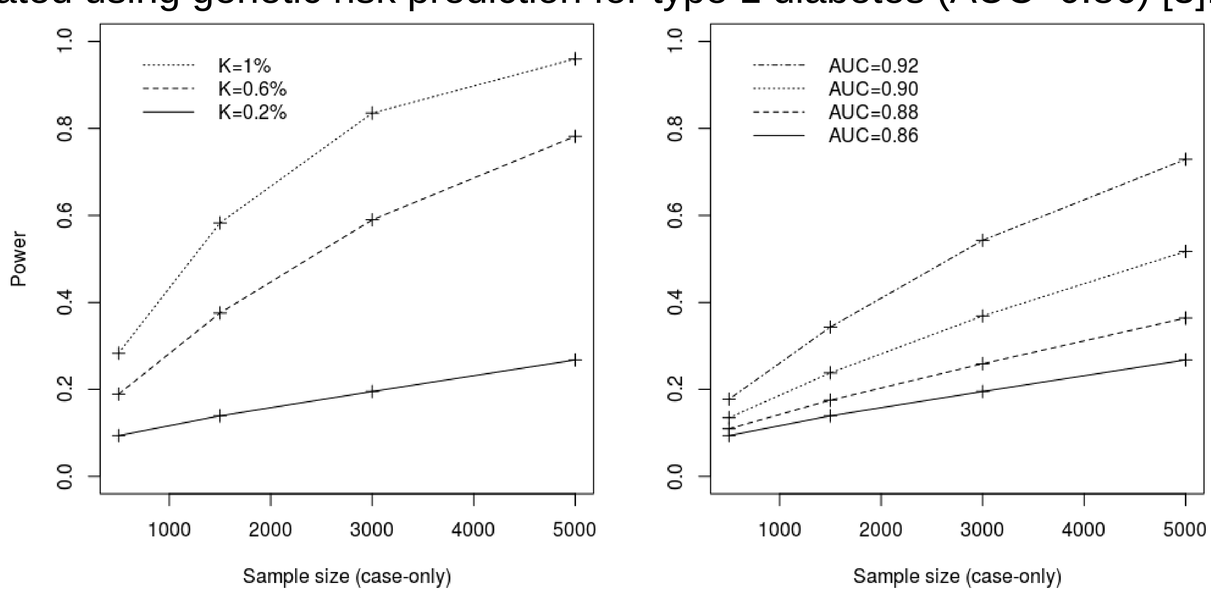
Confirm the association of E and D by finding a negative correlation between E and genetic risk in cases.

Simulation framework

We can quantify the influence of collider bias in different settings by simulating disease occurrence in a population. We use DAC's and the logistic model's assumptions.

Power of DAC

Power increases with sample size, prevalence K and genetic risk prediction accuracy. Simulated using genetic risk prediction for type 1 diabetes (AUC=0.86) [3].



Collider bias in CODGEI

When the rare disease assumption is not respected, **collider bias can change the results of CODGEI. This is the case in Moorman et al. [4]** which applied CODGEI to interactions between BRCA and environment in breast cancer.

References

1 : Cole SR et al. *Int J Epidemiol*. 2010

2 :Piegorsch WW et al. *Stat Med*. 1994

3 :Wei Z et al. *PLOS Genet*. 2009

4 :Moorman PG et al. *Breast Cancer Res Treat*. 2010

Preprint on Biorxiv !