From social determinants to biomarkers: understanding causal pathways

Bianca De Stavola
LSHTM Centre for Statistical Methodology

LSHTM Poverty Symposium
March 17, 2016

Improving health worldwide
www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- Social epidemiology is evolving and adapting to accommodate increasing access to biomarker data

- Social epidemiology is evolving and adapting to accommodate increasing access to biomarker data
- Availability expanding over extended periods of time

- Social epidemiology is evolving and adapting to accommodate increasing access to biomarker data
- Availability expanding over extended periods of time
- This wealth raises several new methodological challenges:

- Social epidemiology is evolving and adapting to accommodate increasing access to biomarker data
- Availability expanding over extended periods of time
- This wealth raises several new methodological challenges:
  (a) How to extend the conceptual framework to include these new dimensions over extended periods of time

- Social epidemiology is evolving and adapting to accommodate increasing access to biomarker data
- Availability expanding over extended periods of time
- This wealth raises several new methodological challenges:
  (a) How to extend the conceptual framework to include these new dimensions over extended periods of time
  (b) How to analyse such complex data

- Social epidemiology is evolving and adapting to accommodate increasing access to biomarker data
- Availability expanding over extended periods of time
- This wealth raises several new methodological challenges:
  (a) How to extend the conceptual framework to include these new dimensions over extended periods of time
  (b) How to analyse such complex data

Recent developments from modern causal inference may be helpful

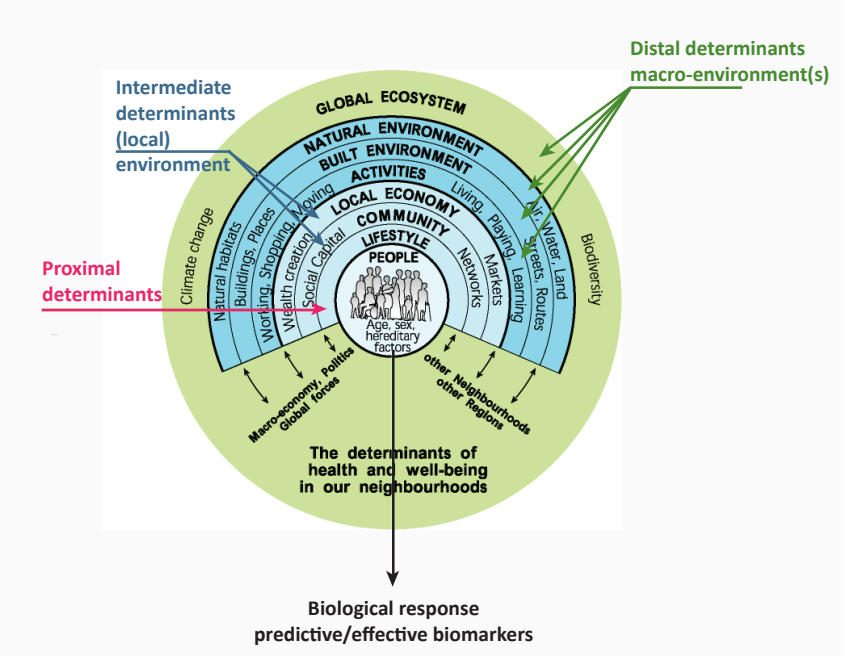
Outline of the talk

Outline of the talk

Inequalities in health

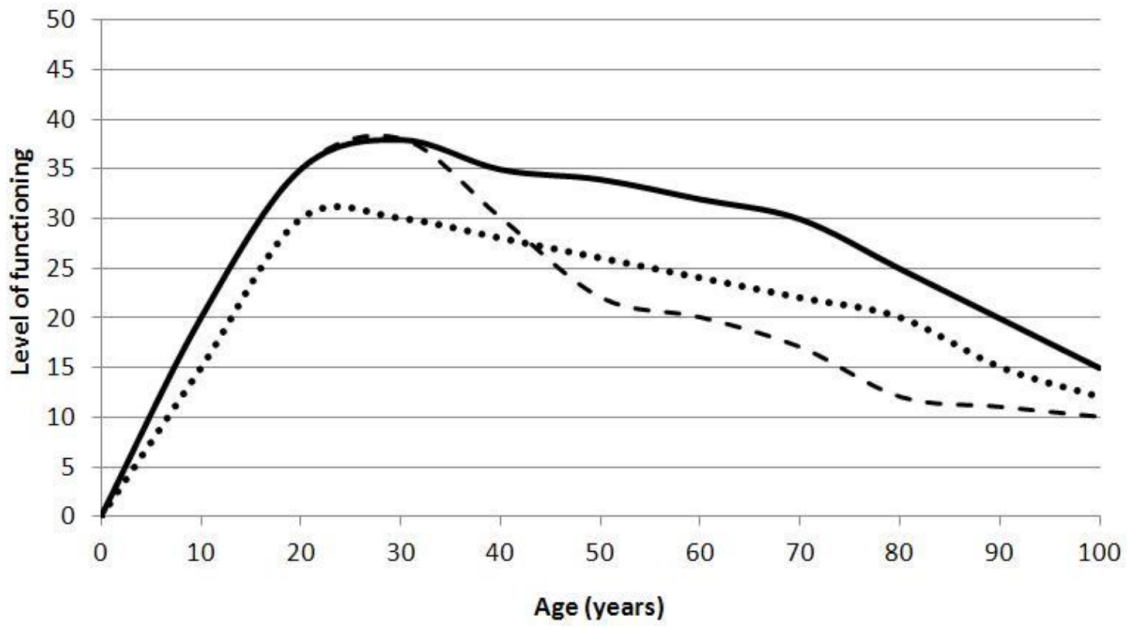
Biomarkers can be viewed as being influenced by exposures that are:

- distal (global environment)
- intermediate (socio-economic)
- proximal (behavioural)



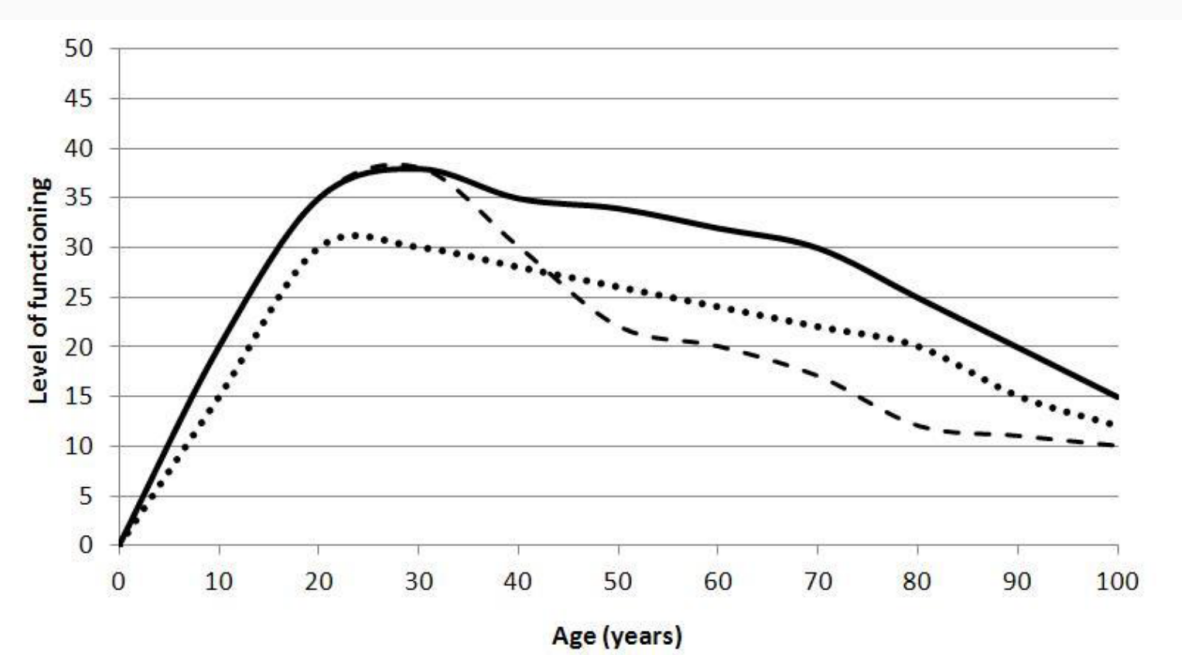
Life course changes

Biomarkers change over the life course:



Life course changes

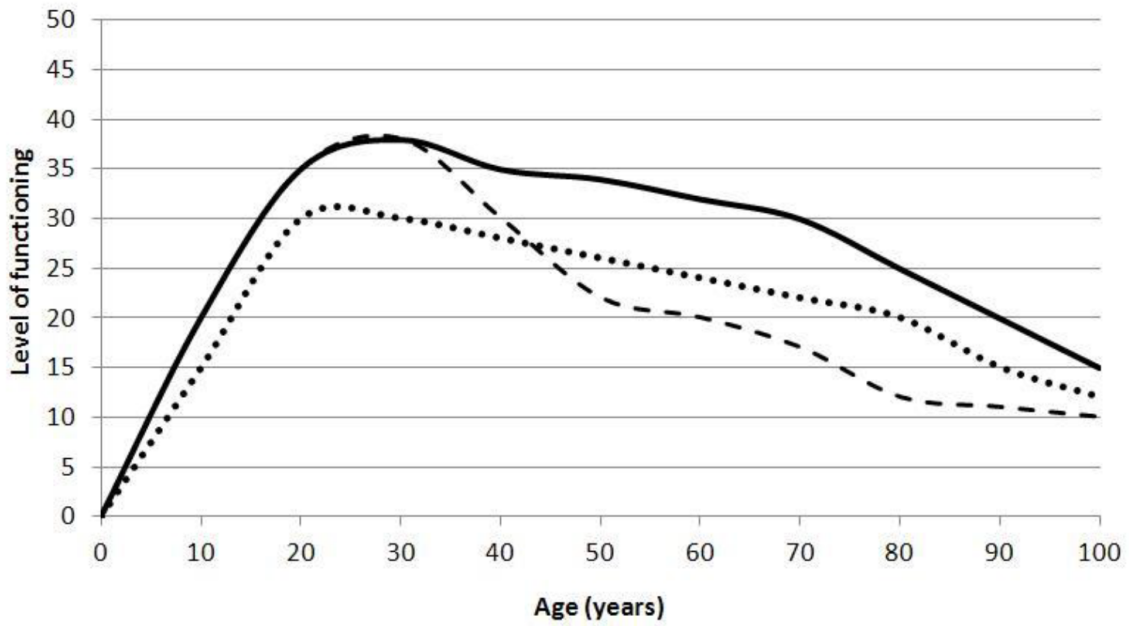
Biomarkers change over the life course:



Their trajectories may be affected by exposures over the life course:

Life course changes

Biomarkers change over the life course:

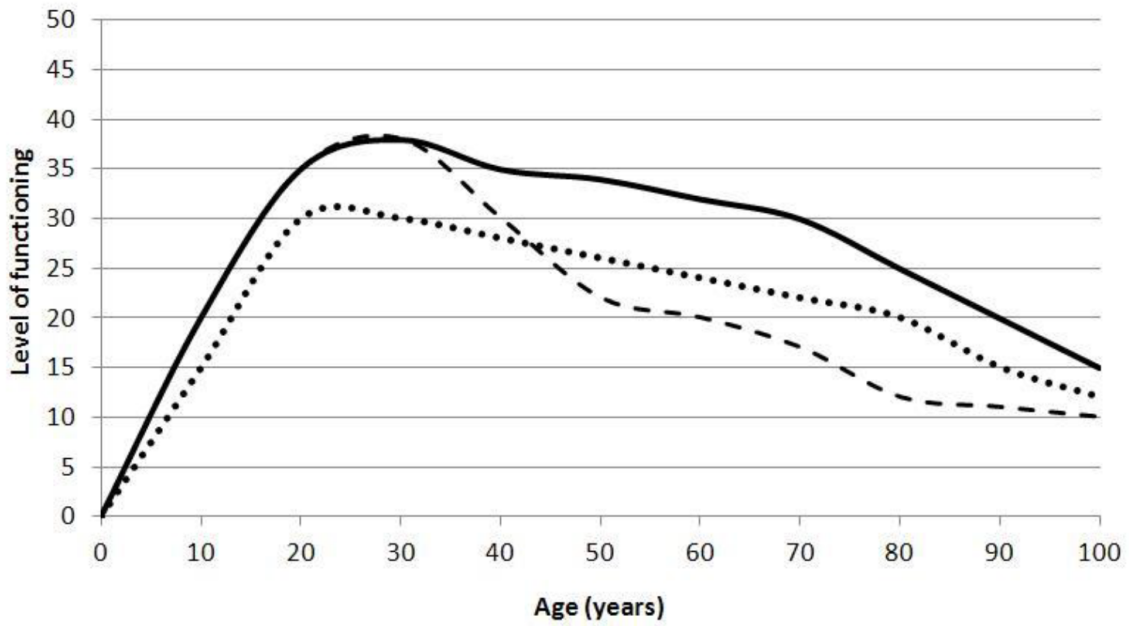


Their trajectories may be affected by exposures over the life course:

[1] cumulatively: → *Cumulative exposure model*

Life course changes

Biomarkers change over the life course:

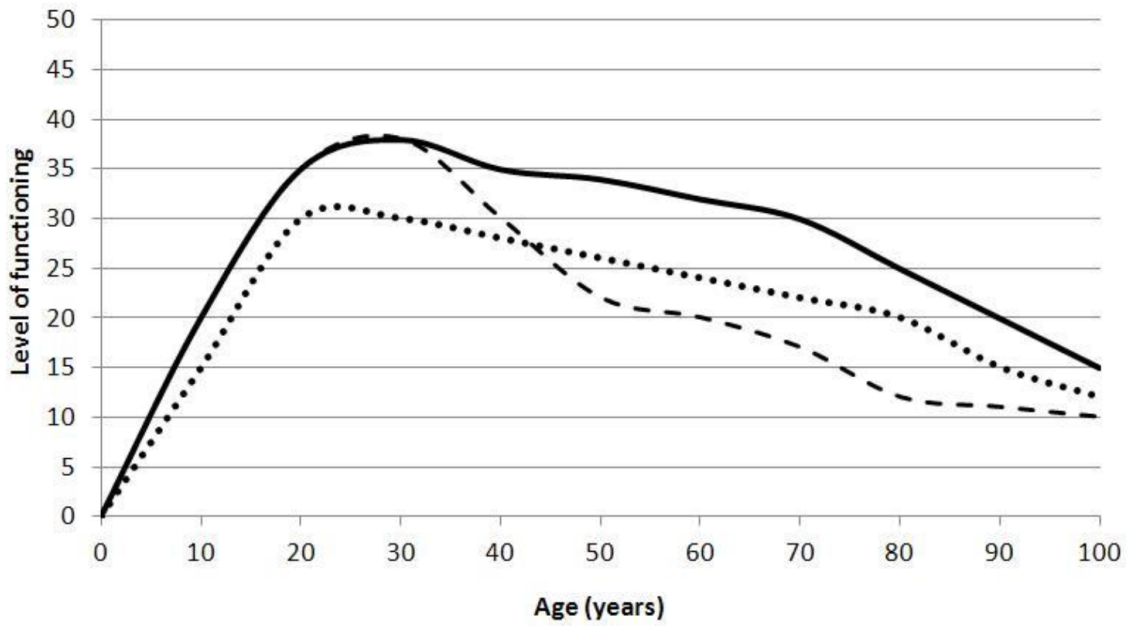


Their trajectories may be affected by exposures over the life course:

[2] only during specific periods: → *Critical period model*

Life course changes

Biomarkers change over the life course:

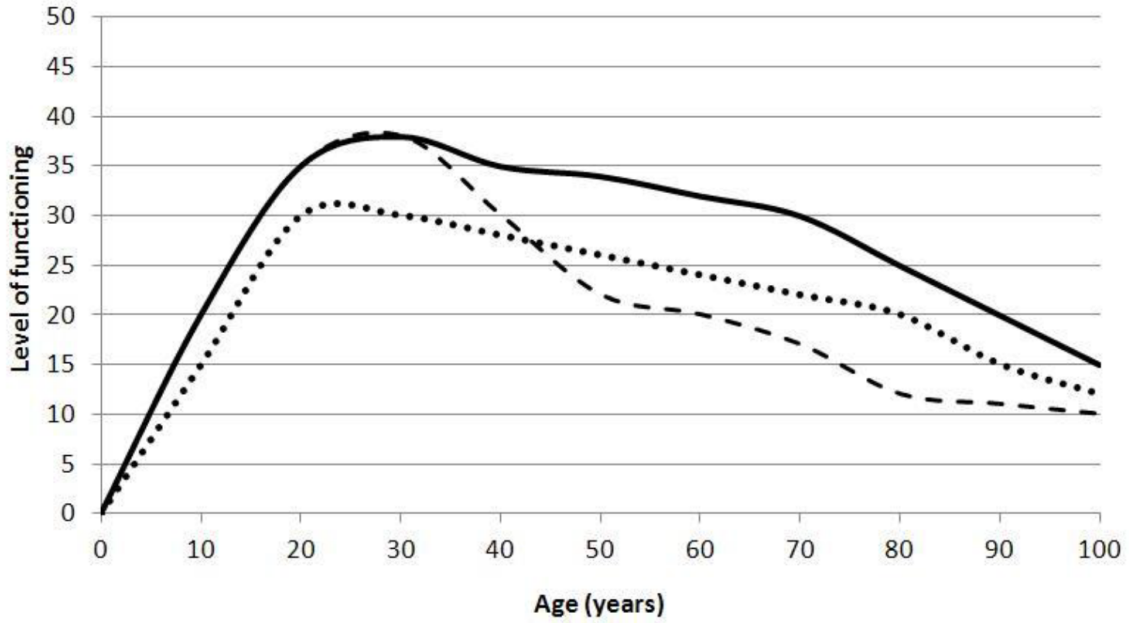


Their trajectories may be affected by exposures over the life course:

[3] especially during specific periods: → *Sensitive period model*

Life course changes

Biomarkers change over the life course:

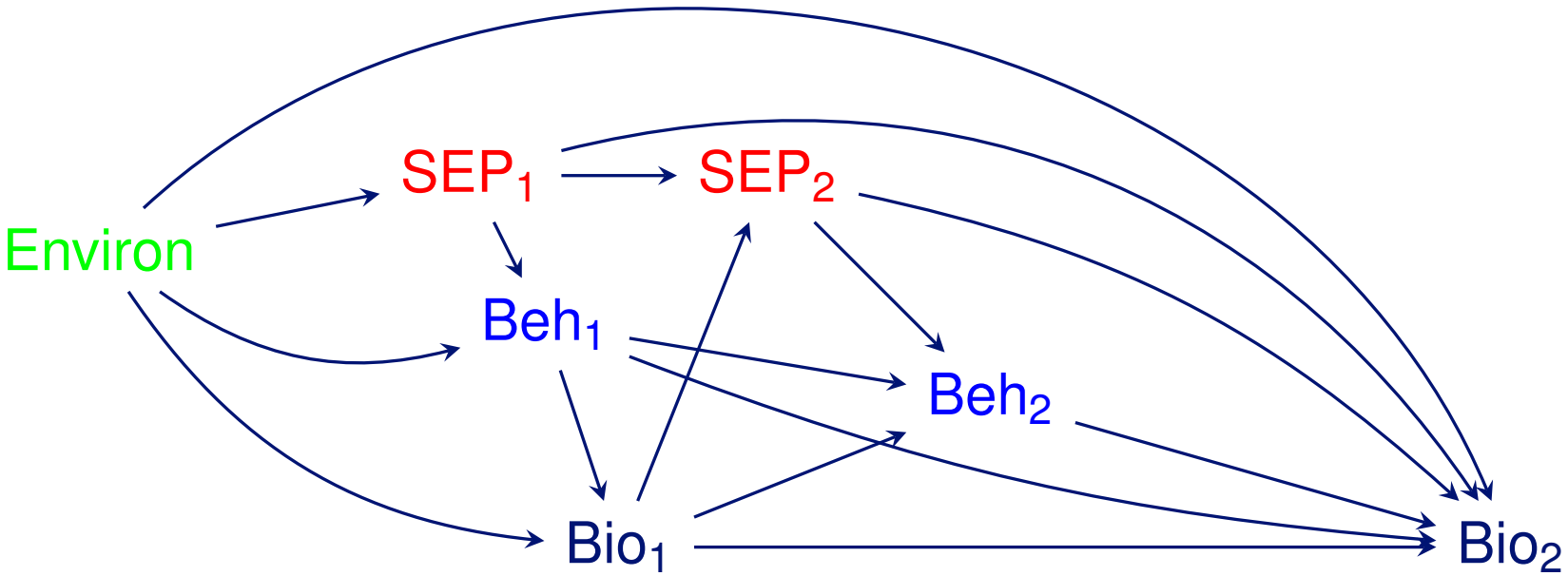


Their trajectories may be affected by exposures over the life course:

[4] synergistically: → *Pathways model*

The conceptual model

The conceptual model could be expanded to include the dimension of time:



Outline of the talk

Questions and causal effects

- We wish to investigate health inequalities

Questions and causal effects

- We wish to investigate health inequalities
- Step from such diagrams to a specific data analysis is complex

Questions and causal effects

- We wish to investigate health inequalities
- Step from such diagrams to a specific data analysis is complex
- Key: clearly formulating the question(s) of interest

Questions and causal effects

- We wish to investigate health inequalities
- Step from such diagrams to a specific data analysis is complex
- Key: clearly formulating the question(s) of interest
- As these are causal investigations, helpful to phrase them as *"How the world would have been had the exposure been different"*

Questions and causal effects

- We wish to investigate health inequalities
- Step from such diagrams to a specific data analysis is complex
- Key: clearly formulating the question(s) of interest
- As these are causal investigations, helpful to phrase them as *"How the world would have been had the exposure been different"*
- This leads to defining potential (counterfactual) outcomes

Questions and causal effects

- We wish to investigate health inequalities
- Step from such diagrams to a specific data analysis is complex
- Key: clearly formulating the question(s) of interest
- As these are causal investigations, helpful to phrase them as *"How the world would have been had the exposure been different"*
- This leads to defining potential (counterfactual) outcomes

Will illustrate how certain causal questions can be formulated using this framework

Potential outcomes

Focussing on socio-economic disadvantage as the exposure of interest and Biomarker at time 2 the outcome, let:

- Y denote the outcome of interest (Bio_2)

Potential outcomes

Focussing on socio-economic disadvantage as the exposure of interest and Biomarker at time 2 the outcome, let:

- Y denote the outcome of interest (Bio_2)
- S_1 and S_2 denote SEP_1 and SEP_1 (binary)

Potential outcomes

Focussing on socio-economic disadvantage as the exposure of interest and Biomarker at time 2 the outcome, let:

- Y denote the outcome of interest (Bio_2)
- S_1 and S_2 denote SEP_1 and SEP_1 (binary)
- $Y(s_1)$ be the *potential outcome*, i.e. the value that Y would take if we were hypothetically to intervene on S_1 and set it to the value s_1

Potential outcomes

Focussing on socio-economic disadvantage as the exposure of interest and Biomarker at time 2 the outcome, let:

- Y denote the outcome of interest (Bio_2)
- S_1 and S_2 denote SEP_1 and SEP_1 (binary)
- $Y(s_1)$ be the *potential outcome*, i.e. the value that Y would take if we were hypothetically to intervene on S_1 and set it to the value s_1
- Similarly for the potential outcome $Y(s_2)$

Potential outcomes

Focussing on socio-economic disadvantage as the exposure of interest and Biomarker at time 2 the outcome, let:

- Y denote the outcome of interest (Bio_2)
- S_1 and S_2 denote SEP_1 and SEP_1 (binary)
- $Y(s_1)$ be the *potential outcome*, i.e. the value that Y would take if we were hypothetically to intervene on S_1 and set it to the value s_1
- Similarly for the potential outcome $Y(s_2)$
- $Y(s_1, s_2)$, the value that Y would take were we to intervene on both S_1 and S_2 and set them to s_1 and s_2

Total causal effect (TCE)

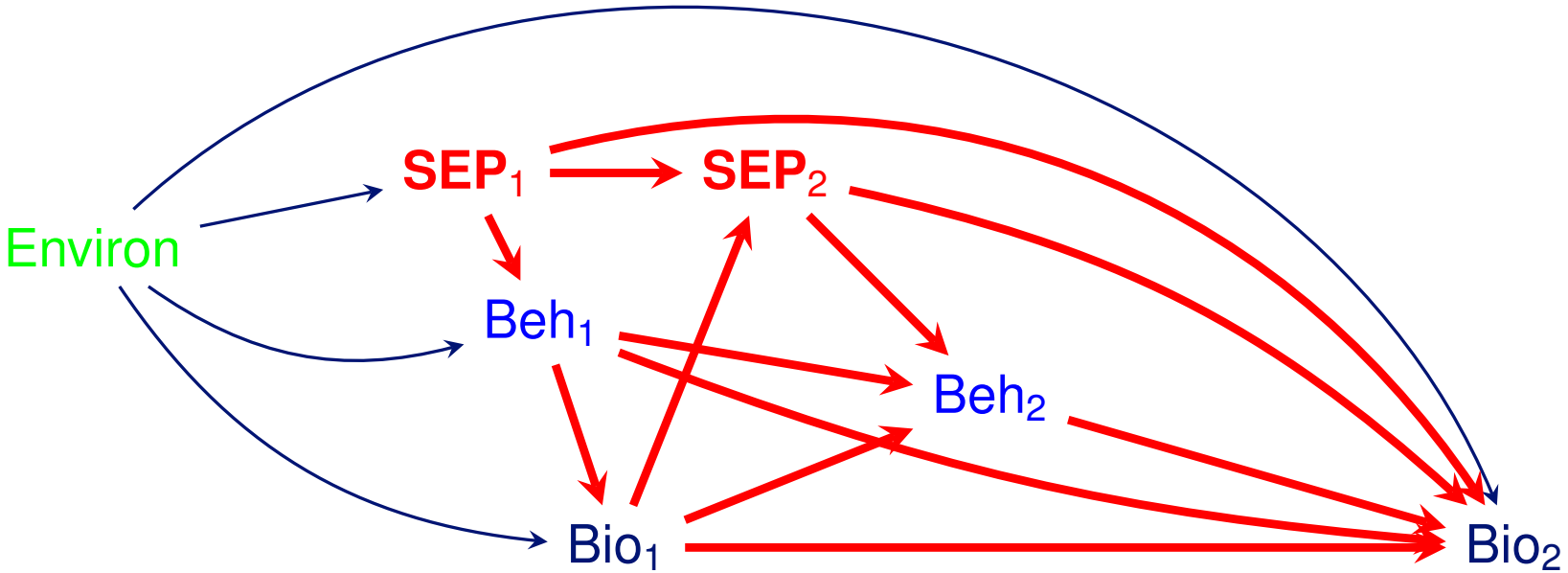
- The total causal effect of S_1 on Y can be expressed as a comparison of aspects of the mean (*i.e.* expectation(E)) of $Y(s_1)$ for different values of s_1,

$$\text{TCE}_1 = E\{Y(s_1 = 1)\} - E\{Y(s_1 = 0)\}$$

Total causal effect (TCE)

- The total causal effect of S_1 on Y can be expressed as a comparison of aspects of the mean (*i.e.* expectation(E)) of $Y(s_1)$ for different values of s_1,

$$\text{TCE}_1 = E\{Y(s_1 = 1)\} - E\{Y(s_1 = 0)\}$$

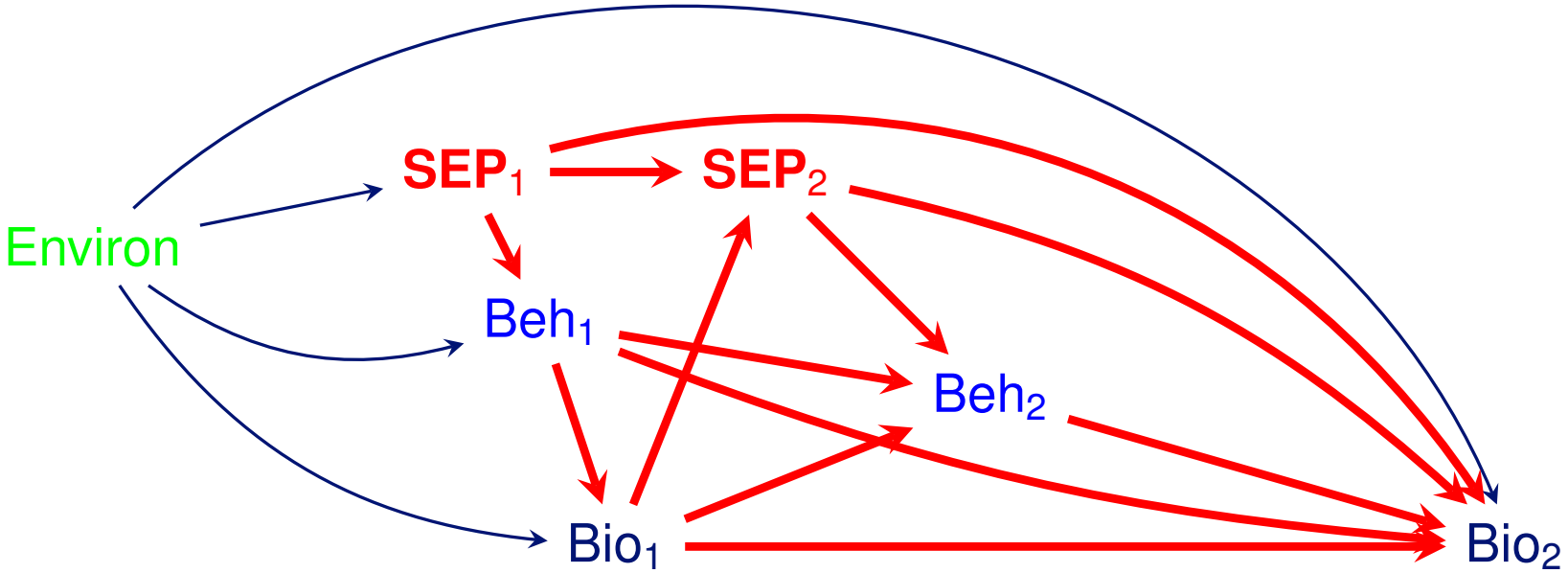


Total causal effect (TCE)

- The total causal effect of S_1 on Y can be expressed as a comparison of aspects of the mean (*i.e.* expectation(E)) of $Y(s_1)$ for different values of s_1,

$$\text{TCE}_1 = E\{Y(s_1 = 1)\} - E\{Y(s_1 = 0)\}$$

- TCE_1 would be of interest to assess the impact of (hypothetically) intervening on SEP_1 and changing its value



Total causal effect (TCE) (cont'd)

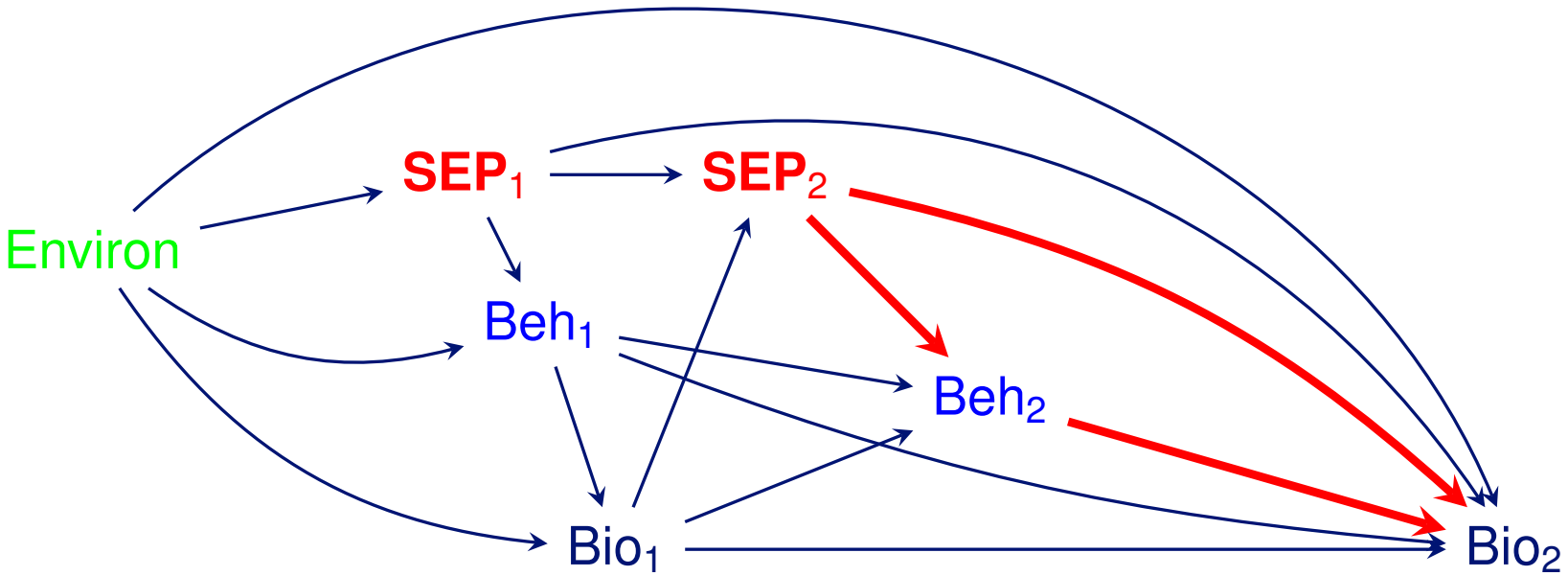
- The total causal effect of S_2 on Y can be expressed as a comparison of the mean of $Y(s_2)$ for different values of s_2,

$$\text{TCE}_2 = E\{Y(s_2 = 1)\} - E\{Y(s_2 = 0)\}$$

Total causal effect (TCE) (cont'd)

- The total causal effect of S_2 on Y can be expressed as a comparison of the mean of $Y(s_2)$ for different values of s_2,

$$\text{TCE}_2 = E\{Y(s_2 = 1)\} - E\{Y(s_2 = 0)\}$$

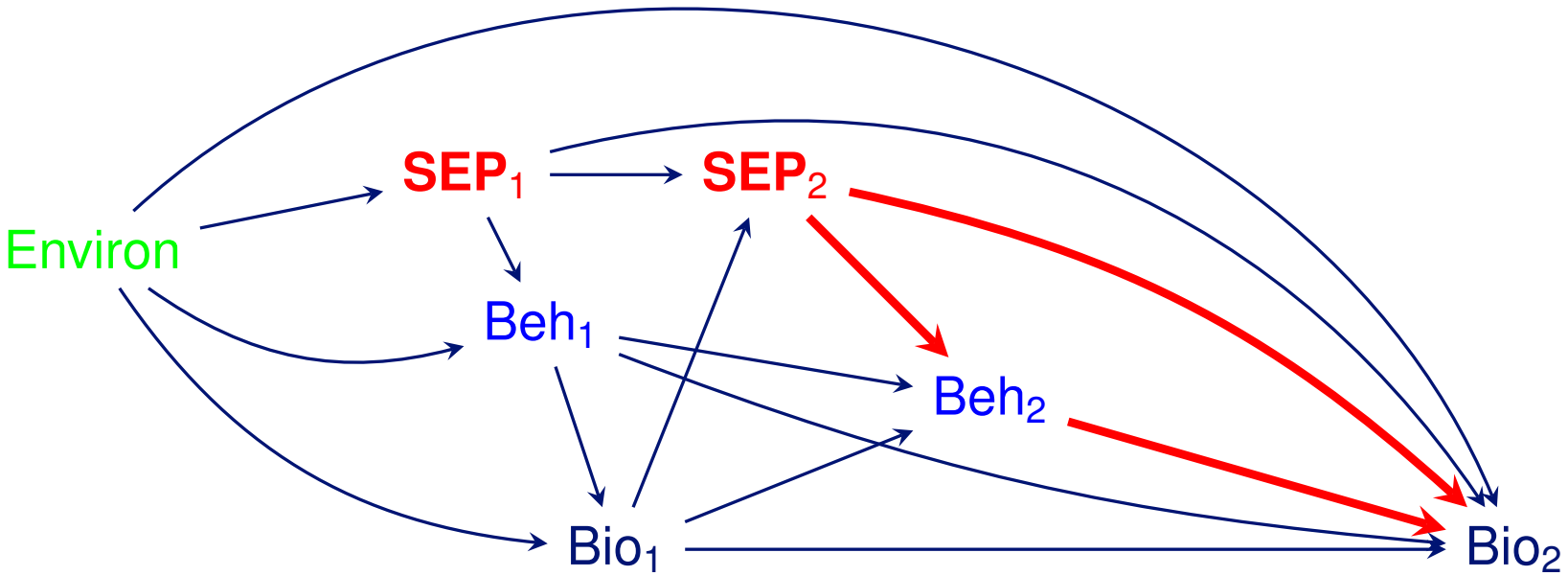


Total causal effect (TCE) (cont'd)

- The total causal effect of S_2 on Y can be expressed as a comparison of the mean of $Y(s_2)$ for different values of s_2,

$$\text{TCE}_2 = E\{Y(s_2 = 1)\} - E\{Y(s_2 = 0)\}$$

- TCE_2 would be of interest to assess the impact of (hypothetically) intervening on SEP_2 and changing its value



Controlled direct causal effect (CDE)

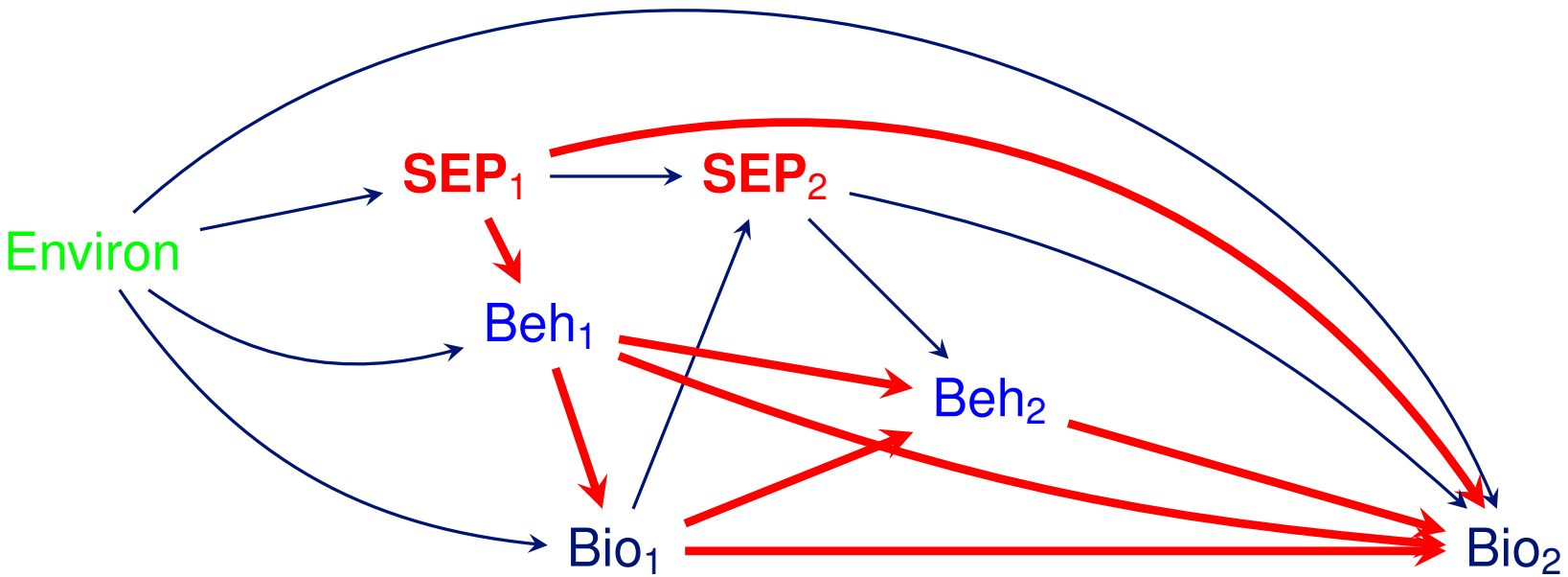
- The CDE of S_1 can be defined in terms of the mean of $Y(s_1, s_2)$ for different values of s_1 but for a fixed value of s_2:

$$\text{CDE}_1(s_2) = E\{Y(s_1 = 1, s_2)\} - E\{Y\{(s_1 = 0, s_2)\}$$

Controlled direct causal effect (CDE)

- The CDE of S_1 can be defined in terms of the mean of $Y(s_1, s_2)$ for different values of s_1 but for a fixed value of s_2:

$$\text{CDE}_1(s_2) = E\{Y(s_1 = 1, s_2)\} - E\{Y\{(s_1 = 0, s_2)\}$$

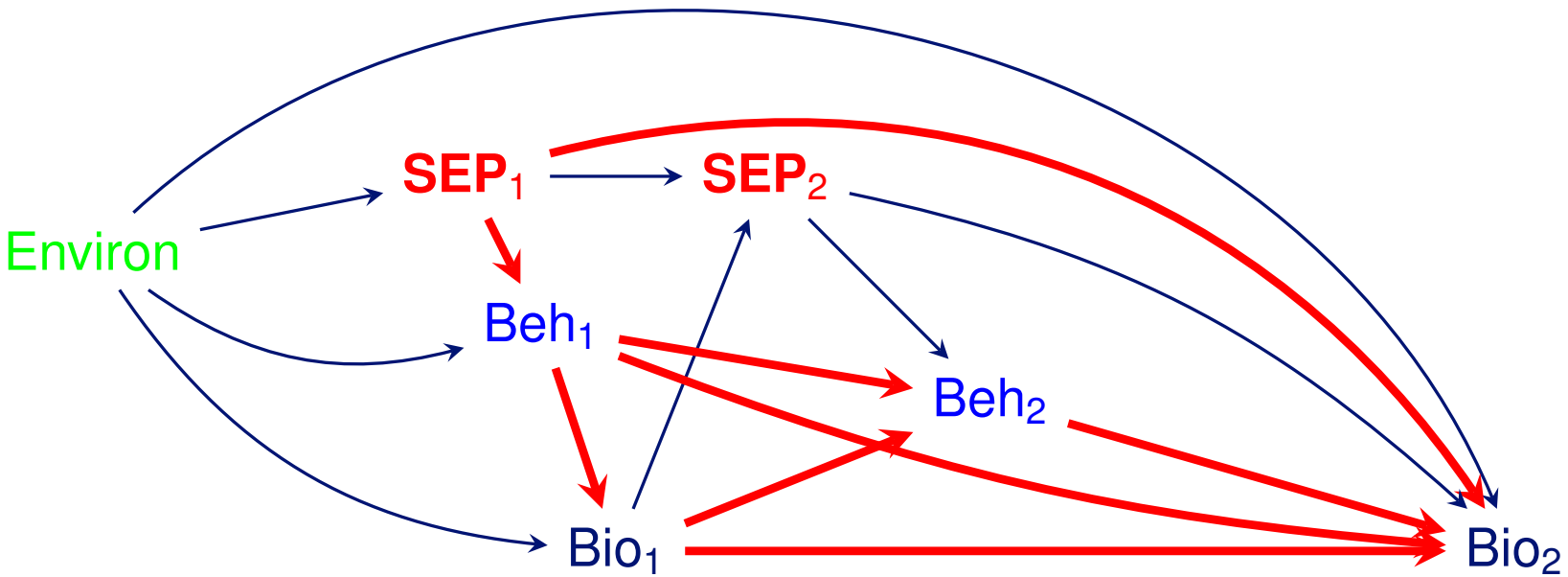


Controlled direct causal effect (CDE)

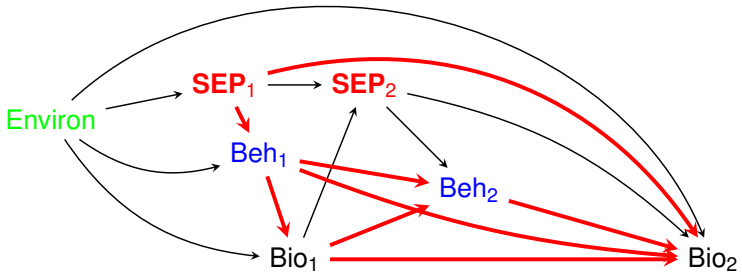
- The CDE of S_1 can be defined in terms of the mean of $Y(s_1, s_2)$ for different values of s_1 but for a fixed value of s_2:

$$\text{CDE}_1(s_2) = E\{Y(s_1 = 1, s_2)\} - E\{Y\{(s_1 = 0, s_2)\}$$

- It represents the direct effect of S_1 that is not mediated by S_2

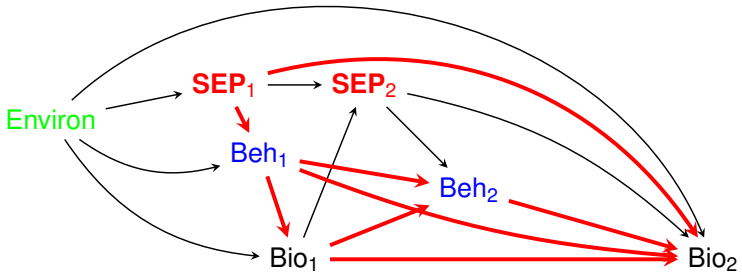


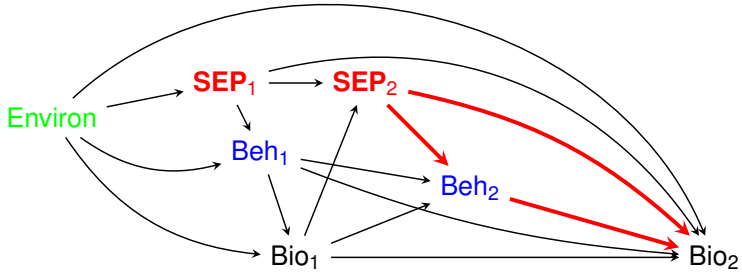
Comparing effects



- If $\text{CDE}_1(s_2)$ varies with s_2 there would be support for the pathways model

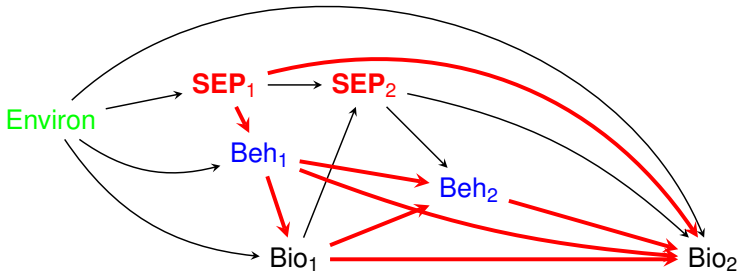
Comparing effects

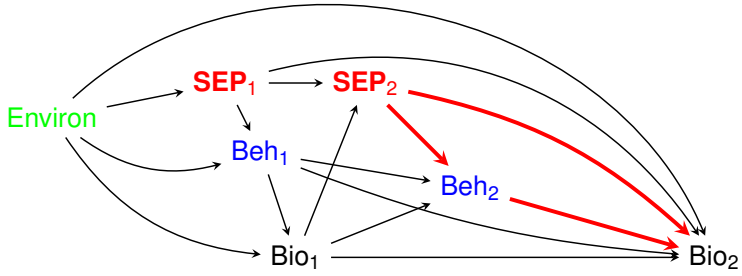




- If $CDE_1(s_2)$ varies with s_2 there would be support for the pathways model
- If CDE_1 does not vary, *i.e.* $CDE_1(s_2) = CDE_1$ and $CDE_1 \sim TCE_2$ there would be support for the cumulative exposure model

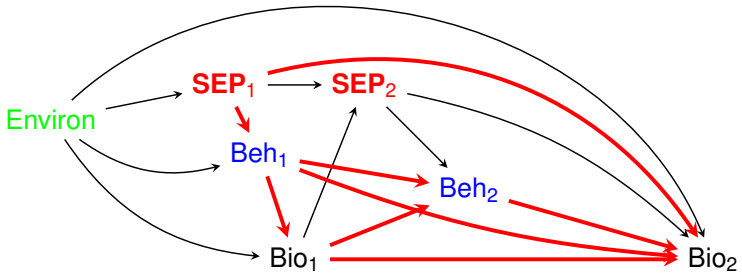
Comparing effects

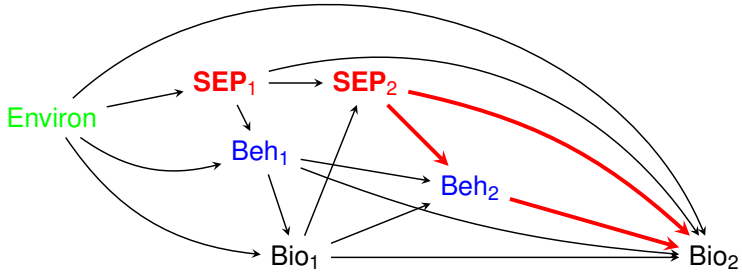




- If $CDE_1(s_2)$ varies with s_2 there would be support for the pathways model
- If CDE_1 does not vary, *i.e.* $CDE_1(s_2) = CDE_1$ and $CDE_1 \sim TCE_2$ there would be support for the cumulative exposure model
- If $CDE_1 \ll TCE_2$ (or $CDE_1 \gg TCE_2$) there would be support for the sensitive period model

Comparing effects





- If $CDE_1(s_2)$ varies with s_2 there would be support for the pathways model
- If CDE_1 does not vary, *i.e.* $CDE_1(s_2) = CDE_1$ and $CDE_1 \sim TCE_2$ there would be support for the cumulative exposure model
- If $CDE_1 \ll TCE_2$ (or $CDE_1 \gg TCE_2$) there would be support for the sensitive period model
- If $CDE_1=0$ or $TCE_2=0$ there would be support for the critical period model

Effect decomposition

Alternatively, we may wish to ask what proportion of the effect of early SEP on the biomarker is mediated by adult SEP.

Effect decomposition

Alternatively, we may wish to ask what proportion of the effect of early SEP on the biomarker is mediated by adult SEP.

- Natural direct and indirect effects (NDE) are relevant to address this

Effect decomposition

Alternatively, we may wish to ask what proportion of the effect of early SEP on the biomarker is mediated by adult SEP.

- Natural direct and indirect effects (NDE) are relevant to address this
- Together they allow the decomposition of the total effect of SEP_1 into the portion that is mediated by SEP_2 and the portion that acts via other pathways.

Effect decomposition

Alternatively, we may wish to ask what proportion of the effect of early SEP on the biomarker is mediated by adult SEP.

- Natural direct and indirect effects (NDE) are relevant to address this
- Together they allow the decomposition of the total effect of SEP_1 into the portion that is mediated by SEP_2 and the portion that acts via other pathways.

Only in very simple settings this decomposition can be achieved by comparing TCE_1 and TCE_2

Comments

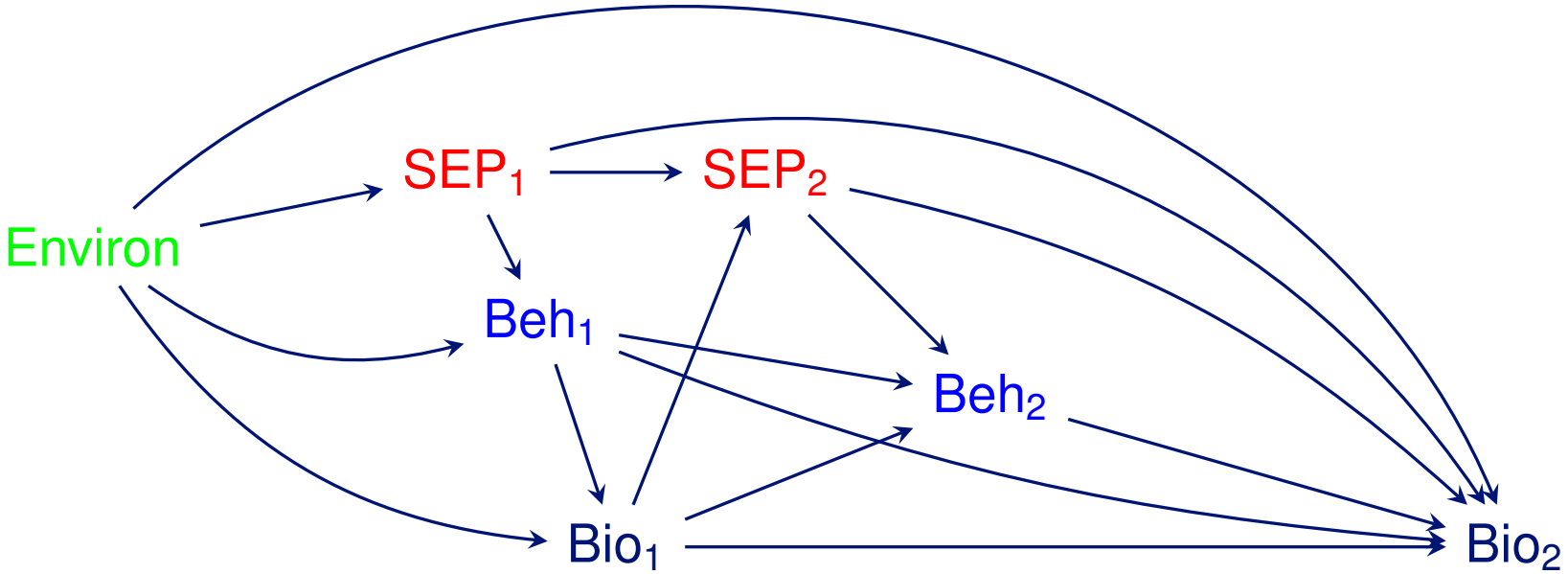
- This is not an exhaustive list of possible causal questions, nor of possible causal parameters
- Given here to illustrate the advantages of adopting the potential outcome framework in order to clarify the question(s) one wishes to address
- Importantly, the chosen question guides the choice of causal parameters to estimate from the data

Outline of the talk

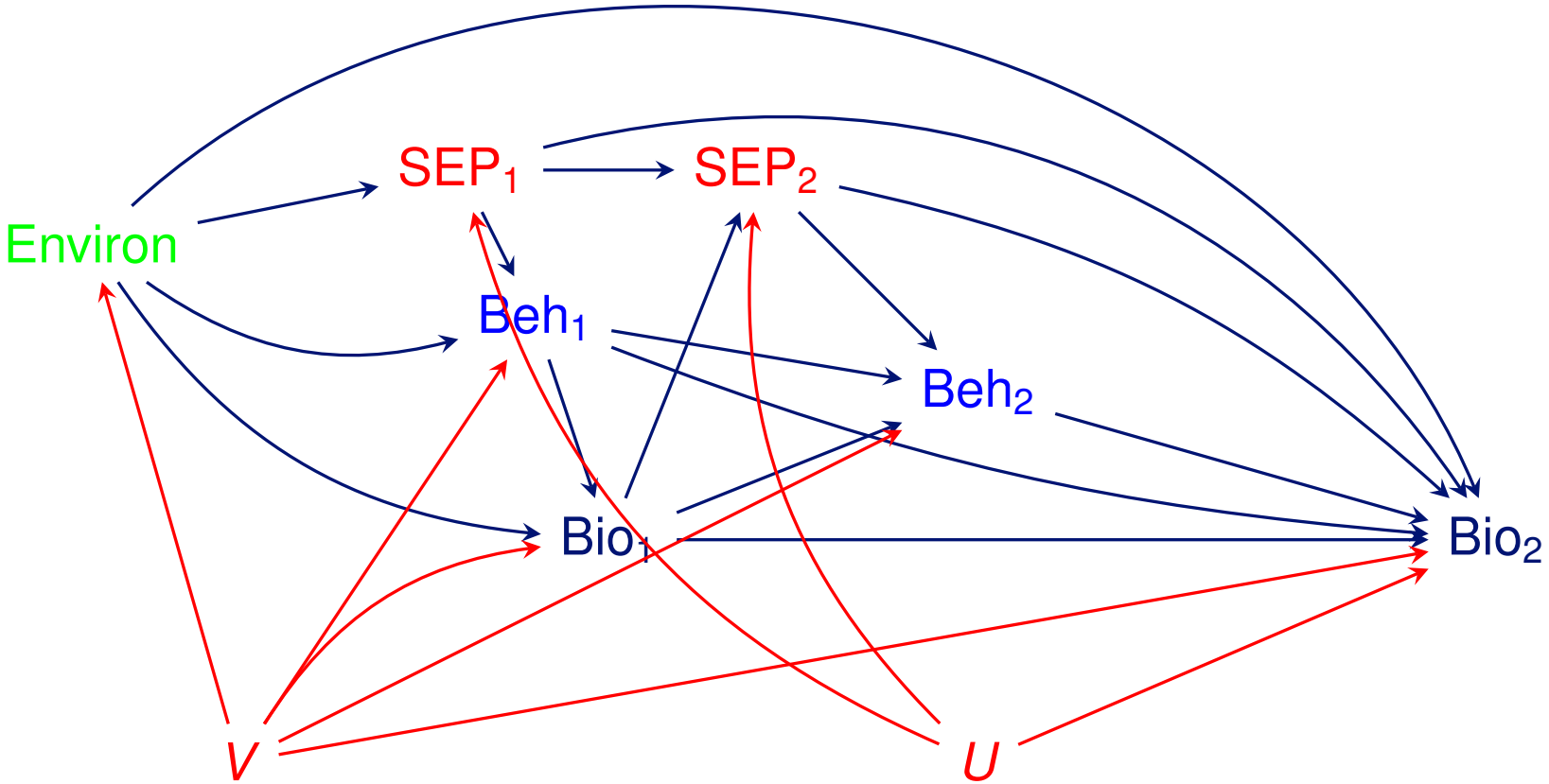
Answering the question

- Given the question, consider assumptions required to estimate the chosen parameter(s) from the data
- Causal DAGs (directed acyclic diagrams) can aid this step
- Unlike the earlier conceptual diagram, causal DAGs are mathematical objects that must obey certain rules
- In particular that any common causes of variables in the diagram must be included, even if unmeasured

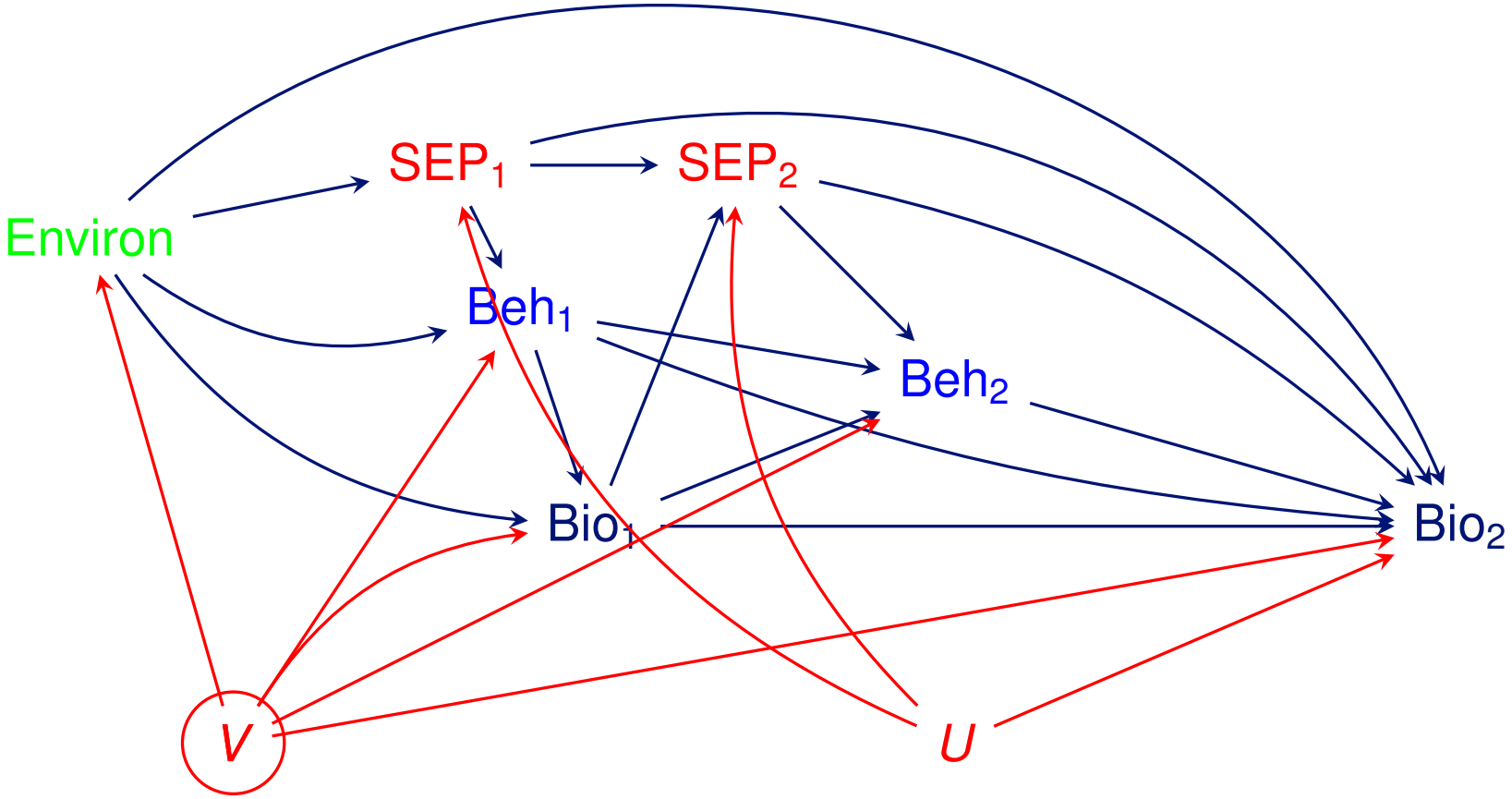
Conceptual and causal diagrams



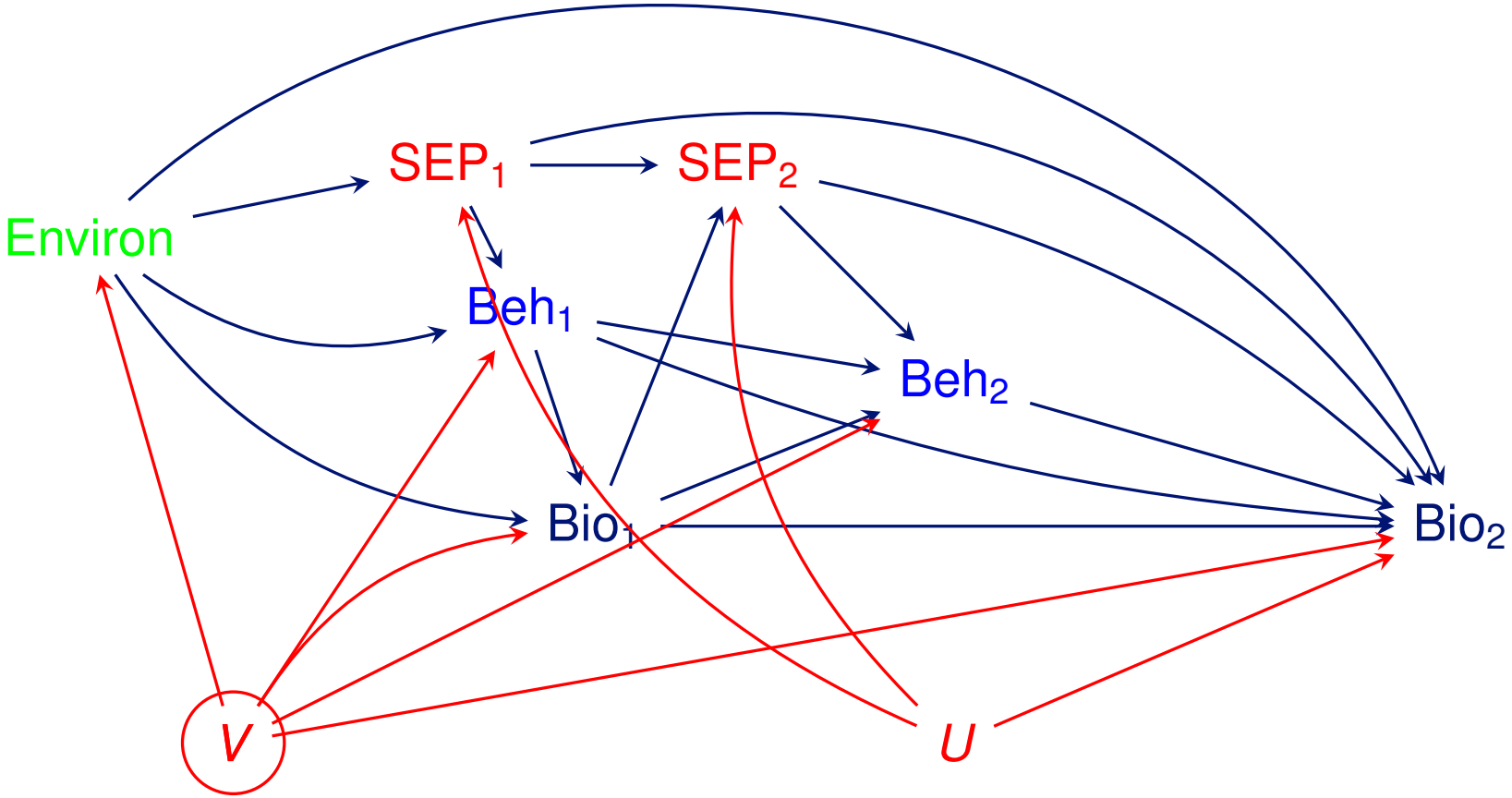
Conceptual and causal diagrams



Conceptual and causal diagrams

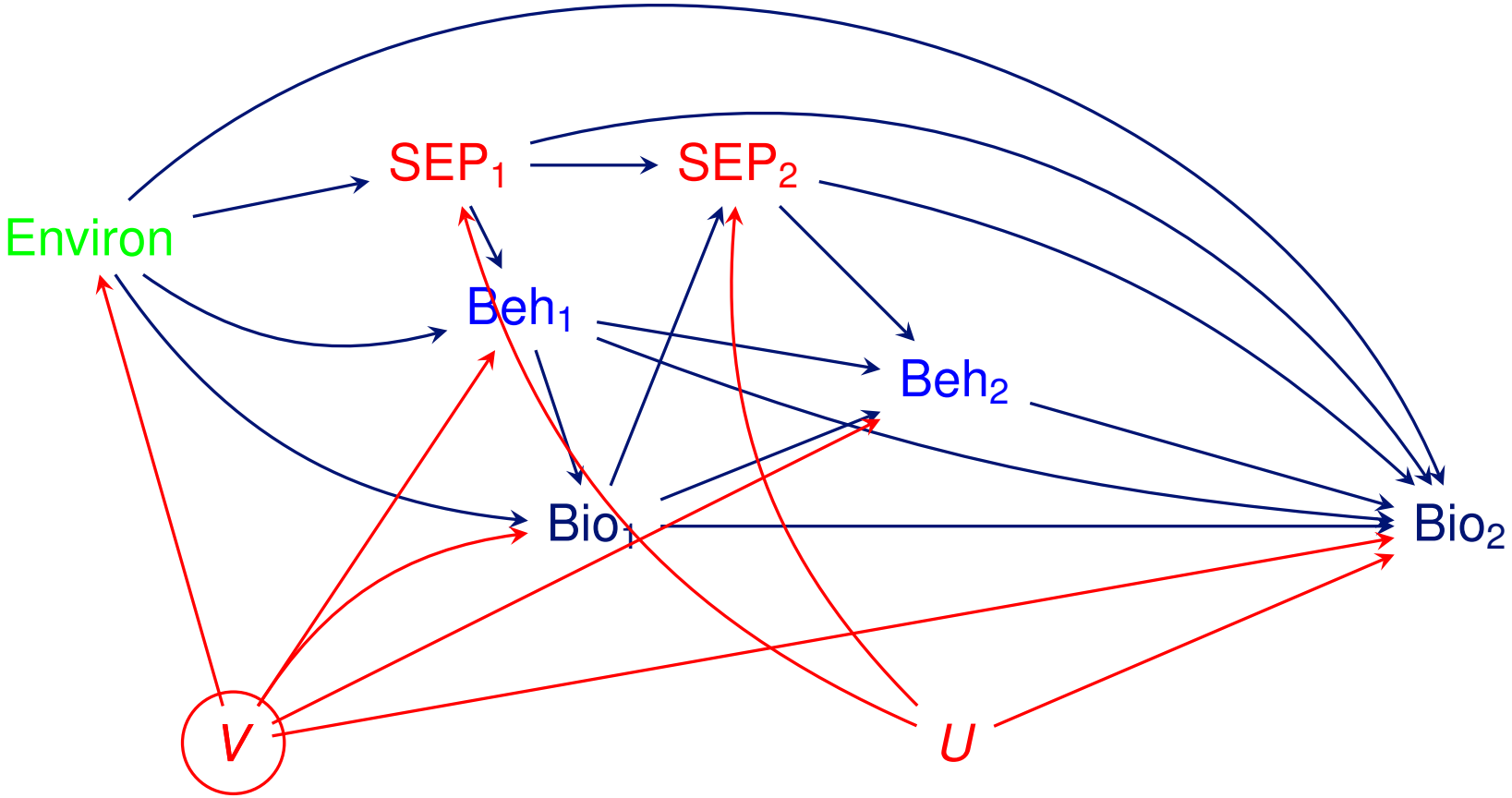


Conceptual and causal diagrams



The DAG can be interrogated to find out whether the chosen parameter can be estimated from the available data and also whether traditional regression methods would suffice.

Conceptual and causal diagrams



The DAG can be interrogated to find out whether the chosen

Recent developments from modern causal inference may be helpful

Outline of the talk

Summary

- The availability of richer data to investigate health inequalities raises daunting methodological challenges

Summary

- The availability of richer data to investigate health inequalities raises daunting methodological challenges
- Concepts (and methods) from the field of causal inference can be useful to formalise the questions in a clear and unambiguous manner

Summary

- The availability of richer data to investigate health inequalities raises daunting methodological challenges
- Concepts (and methods) from the field of causal inference can be useful to formalise the questions in a clear and unambiguous manner
- They are also extremely useful for identifying the appropriate estimation methods

Reference

De Stavola BL and Daniel R
"Commentary: Incorporating concepts and methods from causal inference into life course epidemiology"
Int J Epi, 2016 (in press)