

Directed acyclic graphs in Neighborhood and Health research (Social Epidemiology)

Basile Chaix

Inserm, France

Etsuji Suzuki

Okayama University, Japan



Inserm

Institut national
de la santé et de la recherche médicale



Inference in n'hood & health research

N'hood (neighborhood) and health research has to deal with challenging difficulties:

- N'hood factors are distal causes of health: significant causal distance
- Complex causal chains with variables at multiple levels: cross-level mediating mechanisms
- Selective migration as a source of confounding
- Ecological processes generating interdependence among the n'hood exposures
- Feedback loops and reciprocal interactions

Inference in n'hood & health research

Randomization in n'hood and health research?

(i) of interventions to neighborhoods

(ii) of individuals to neighborhoods

(i) Randomized community trials

- in a limited number of n'hoods → generalizability?
- only applicable to a restricted range of n'hood exposures
- which of the multiple components is influential?

(ii) Residential relocation programs

- in a limited number of n'hoods
- unnatural scenario + ethical issues

→ Interventions often only representative of themselves

→ Observational studies remain a key approach

DAGs: GENERAL RATIONALE

Common problems:

- Imprecise identification of research hypotheses
- Inappropriate selection of adjustment variables

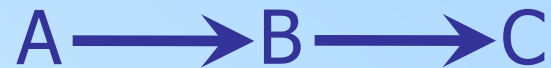
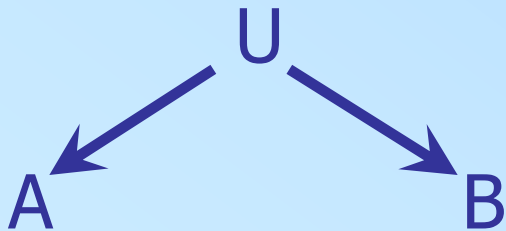
Directed Acyclic Graphs (DAGs) allow to:

- graphically encode *a priori* assumptions about causal relations between exposure, outcome, and covariates (before data analysis)
 - identify appropriate analytical strategies
- depict alternative sets of causal structures that could give rise to observed associations (after or during data analysis)

DAGs: STRUCTURE AND BASIC RULES

Causal DAGs are composed of:

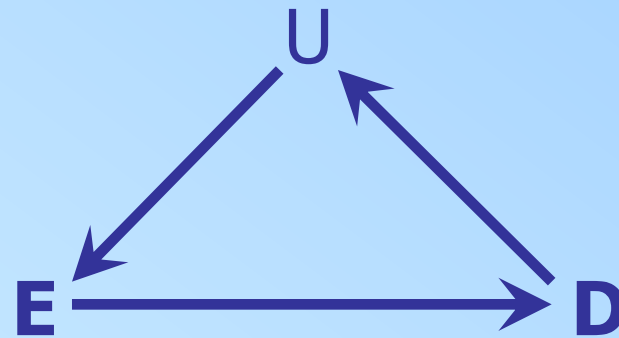
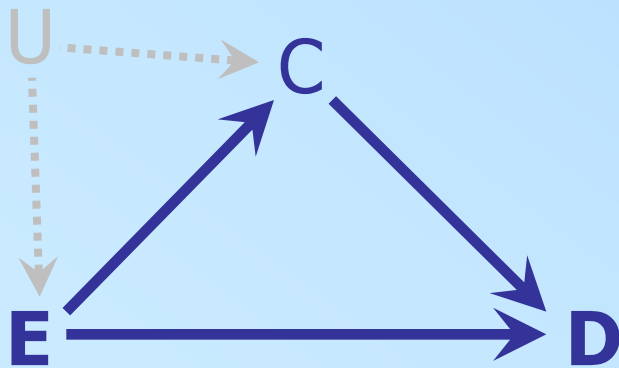
- nodes: representing (un)measured variables
 - In our case: individuals as units of analysis
(n'hood variables reflect n'hood exposures)
- directed arrows (or edges) between variables
(most often single-headed)
 - Arrows can be interpreted as direct causal effects
 - Sequence of directed arrows: indirect effect



DAGs: STRUCTURE AND BASIC RULES

Formal rules and assumptions of DAGs:

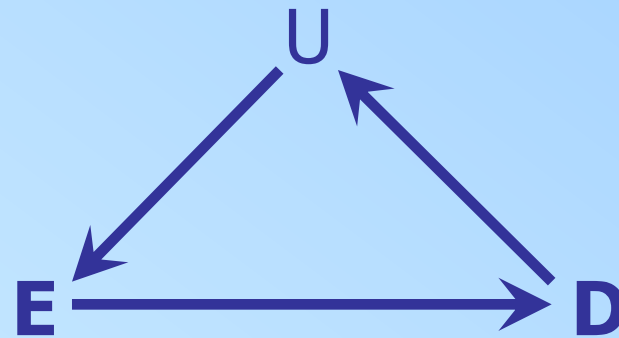
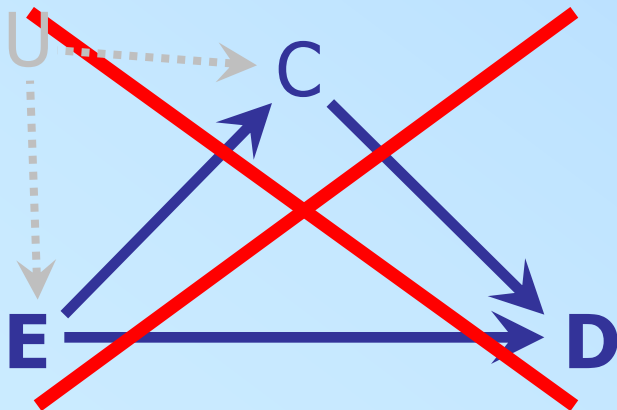
- If two variables in the DAG share a common cause (including an unmeasured one), it has to be reported.
- **Acyclic**: a variable cannot cause itself (directly or indirectly)



DAGs: STRUCTURE AND BASIC RULES

Formal rules and assumptions of DAGs:

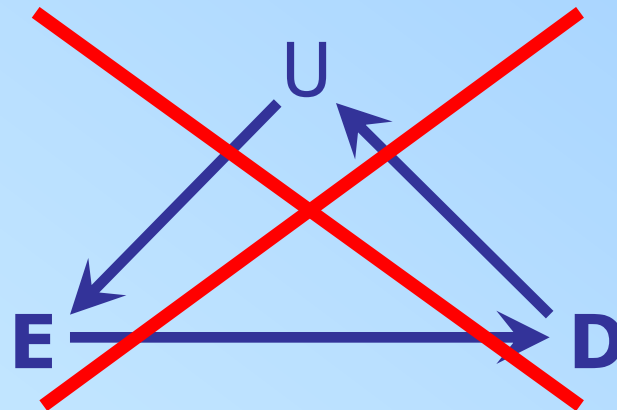
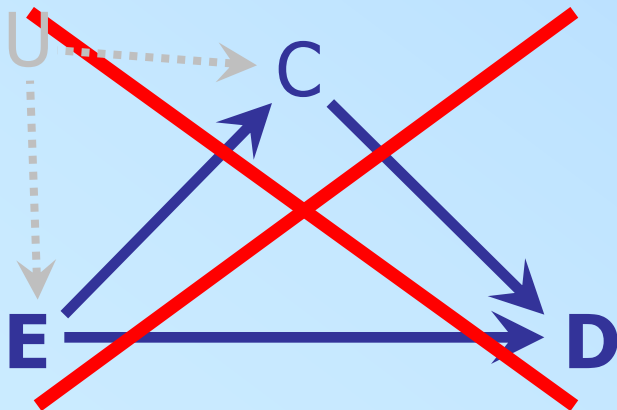
- If two variables in the DAG share a common cause (including an unmeasured one), it has to be reported.
- Acyclic: a variable cannot cause itself (directly or indirectly)



DAGs: STRUCTURE AND BASIC RULES

Formal rules and assumptions of DAGs:

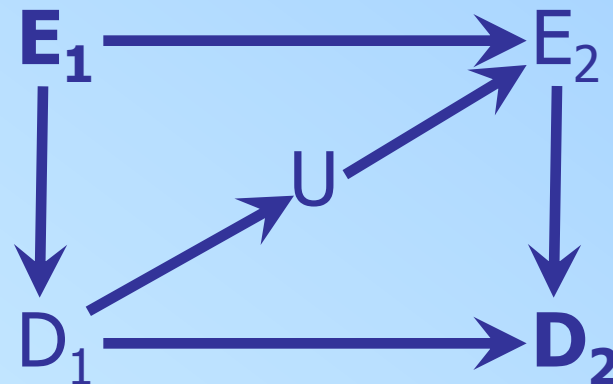
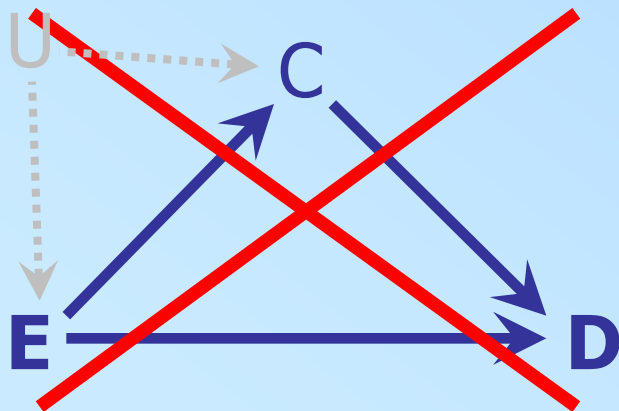
- If two variables in the DAG share a common cause (including an unmeasured one), it has to be reported.
- Acyclic: a variable cannot cause itself (directly or indirectly)



DAGs: STRUCTURE AND BASIC RULES

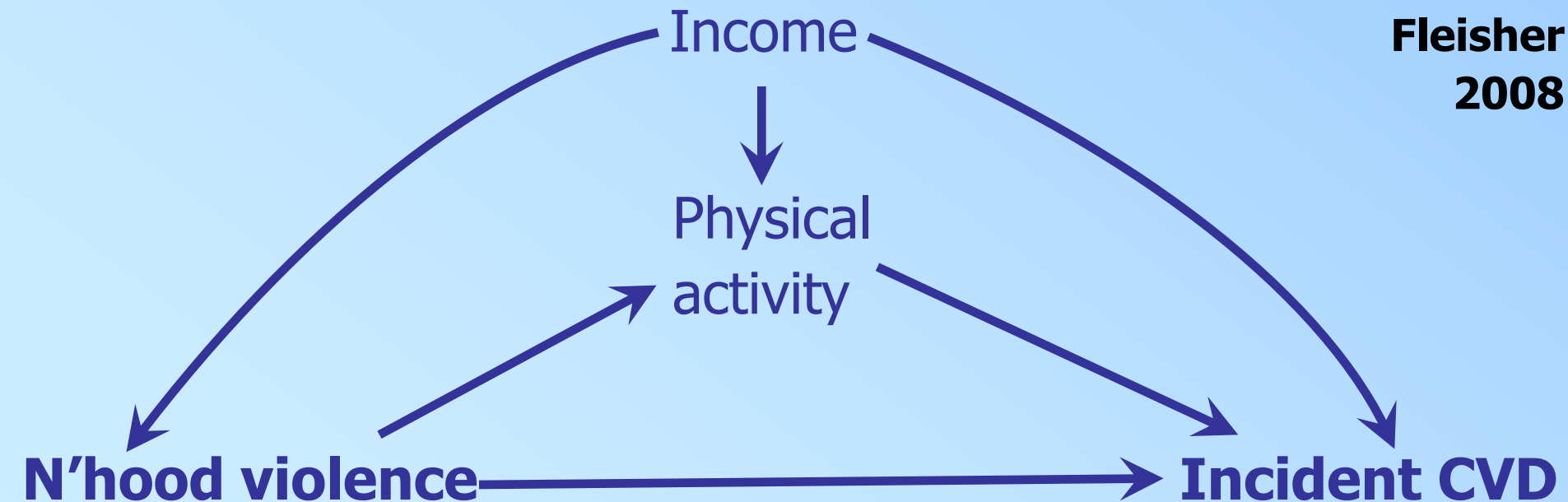
Formal rules and assumptions of DAGs:

- If two variables in the DAG share a common cause (including an unmeasured one), it has to be reported.
- Acyclic: a variable cannot cause itself (directly or indirectly)



DAGs: TERMINOLOGY

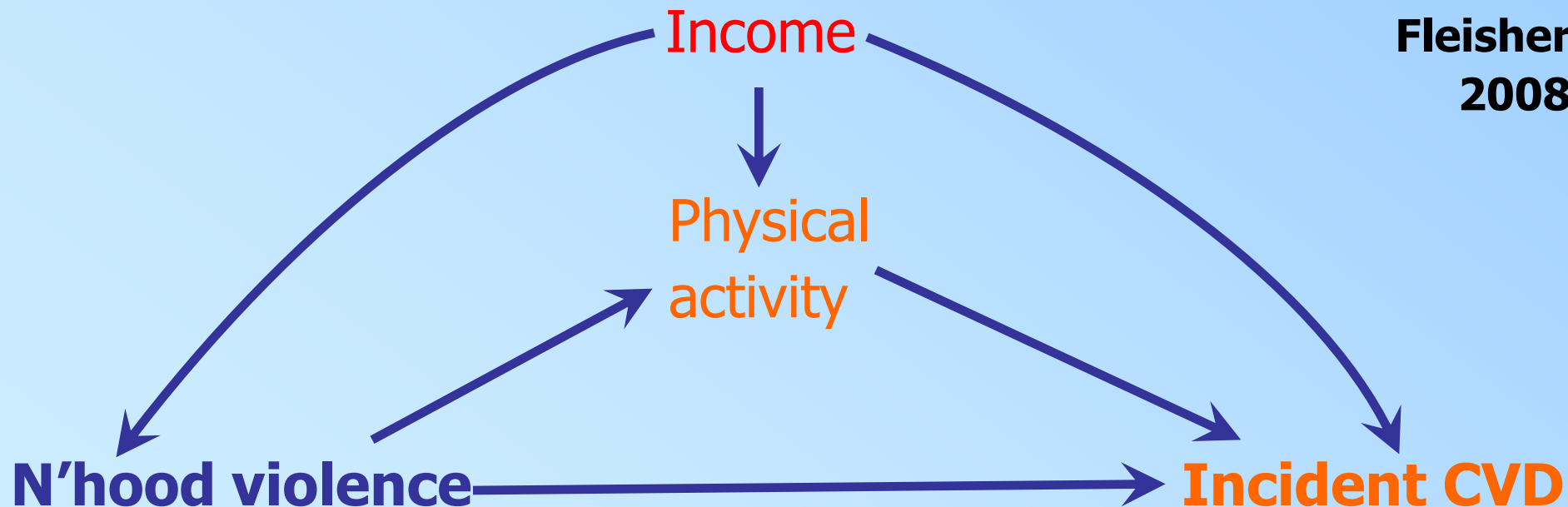
- A **child** of a variable / a **parent** of a variable
- A **descendant** of a variable / an **ancestor** of a variable
- **A path**: a series of lines connecting two variables, regardless of arrowhead direction



DAGs: TERMINOLOGY

- A **child** of a variable / a **parent** of a variable
- A **descendant** of a variable / an **ancestor** of a variable
- **A path**: a series of lines connecting two variables, regardless of arrowhead direction

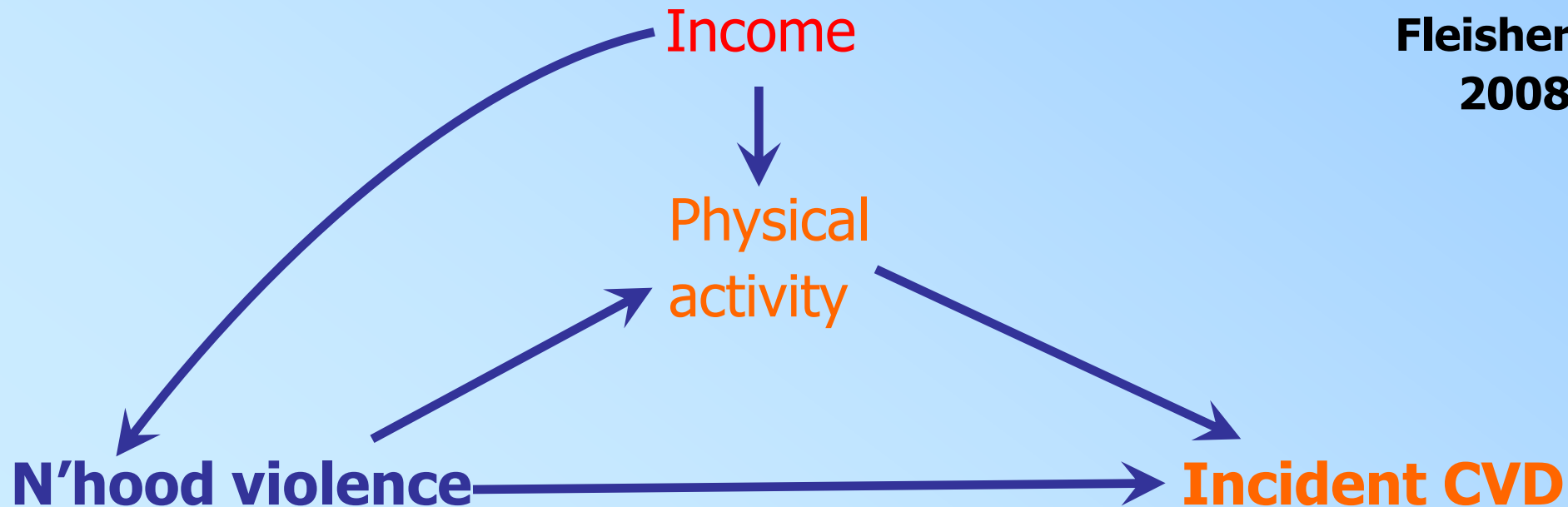
**Fleisher
2008**



DAGs: TERMINOLOGY

- A **child** of a variable / a **parent** of a variable
- A **descendant** of a variable / an **ancestor** of a variable
- **A path**: a series of lines connecting two variables, regardless of arrowhead direction

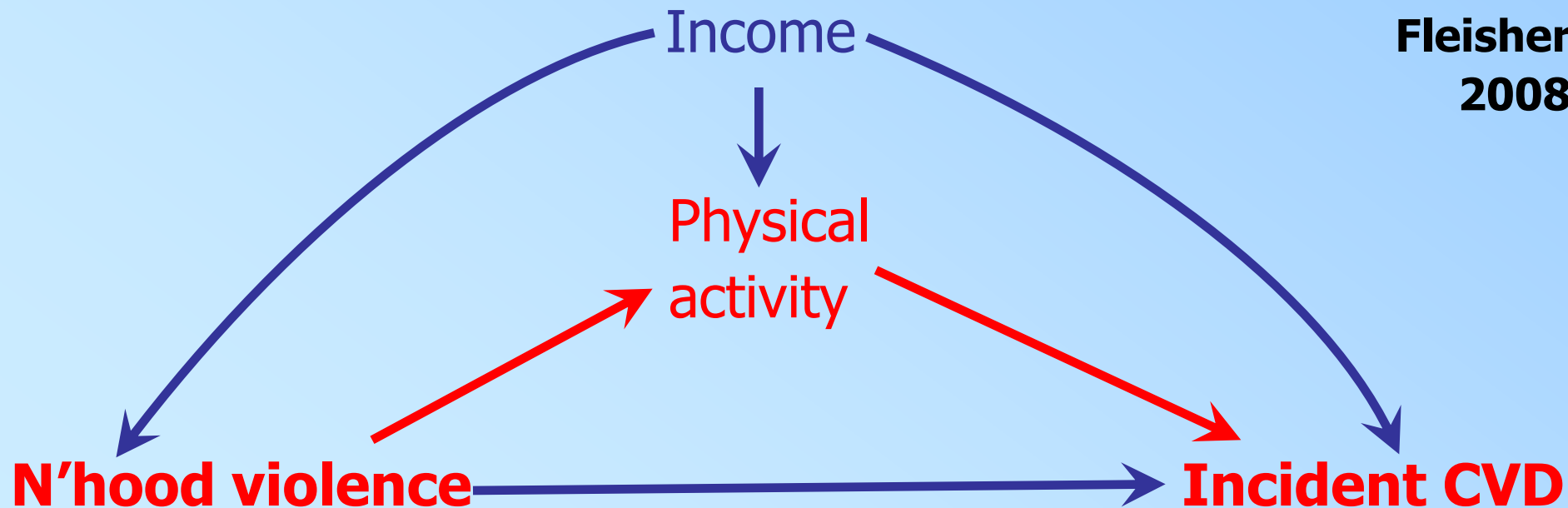
**Fleisher
2008**



DAGs: TERMINOLOGY

- A **child** of a variable / a **parent** of a variable
- A **descendant** of a variable / an **ancestor** of a variable
- **A path**: a series of lines connecting two variables, regardless of arrowhead direction

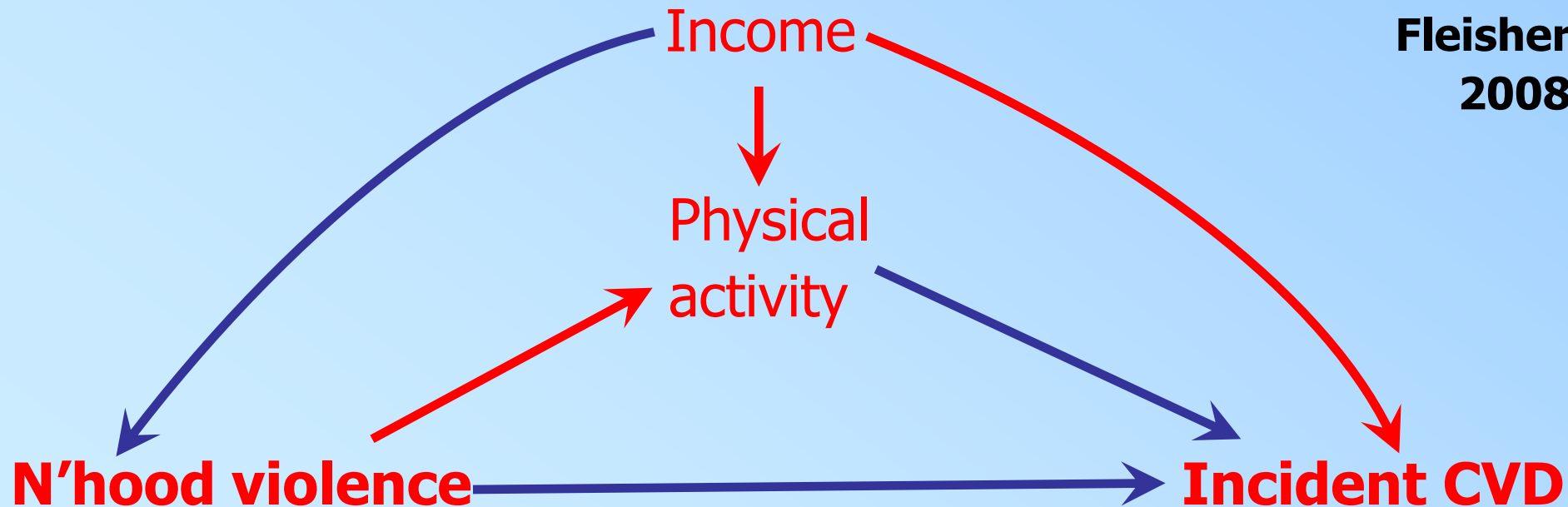
**Fleisher
2008**



DAGs: TERMINOLOGY

- A **child** of a variable / a **parent** of a variable
- A **descendant** of a variable / an **ancestor** of a variable
- A **path**: a series of lines connecting two variables, regardless of arrowhead direction

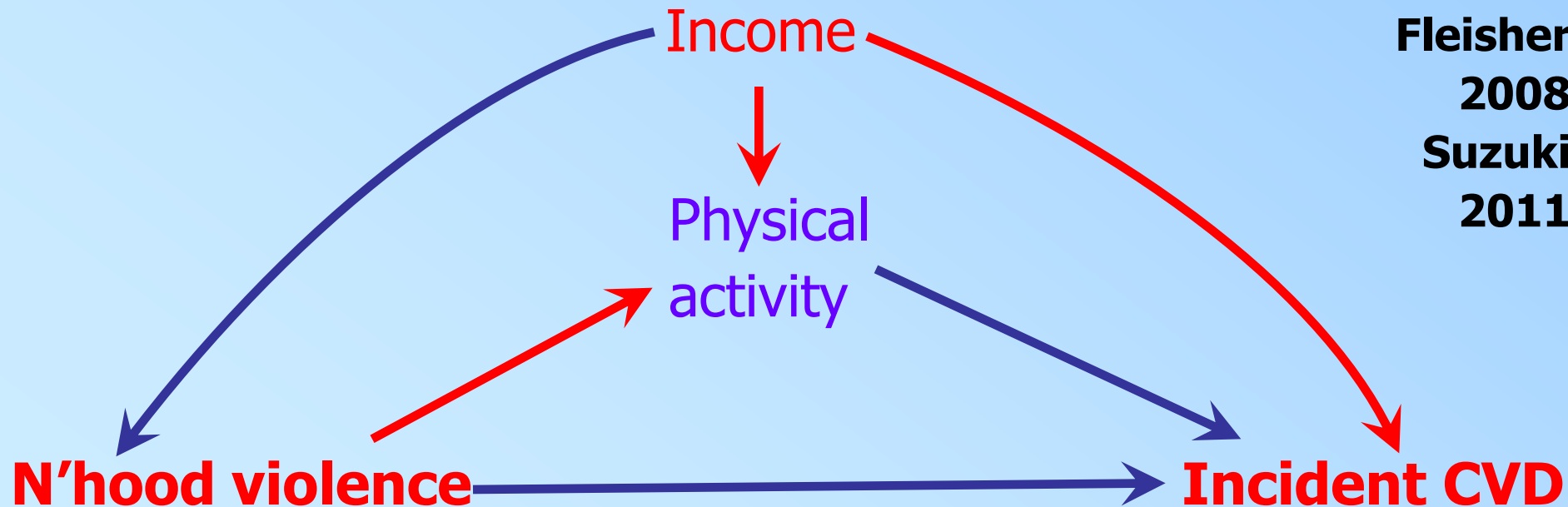
**Fleisher
2008**



DAGs: TERMINOLOGY

- A **collider** on a **path**: a variable with two arrows into it (common effect): where two arrows “collide”
- **Unblocked path**: sequence of arrows connecting two variables that does not contain a collider

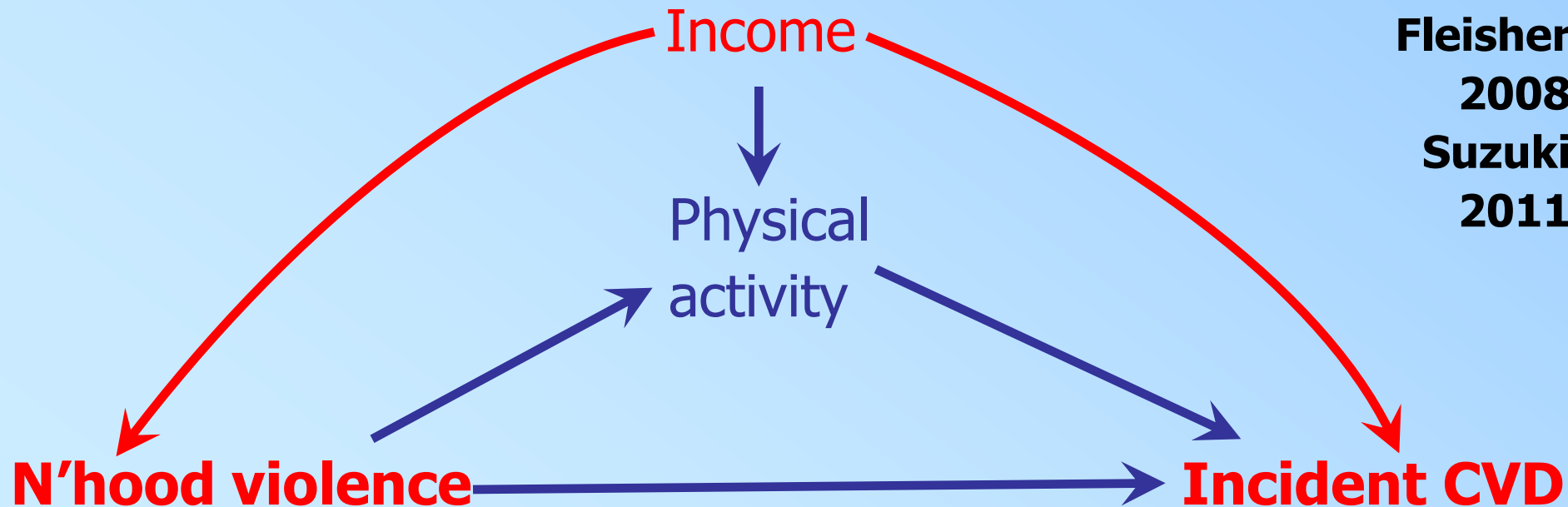
**Fleisher
2008
Suzuki
2011**



DAGs: TERMINOLOGY

- **A collider on a path:** a variable with two arrows into it (common effect): where two arrows “collide”
- **Unblocked path:** sequence of arrows connecting two variables that does not contain a collider

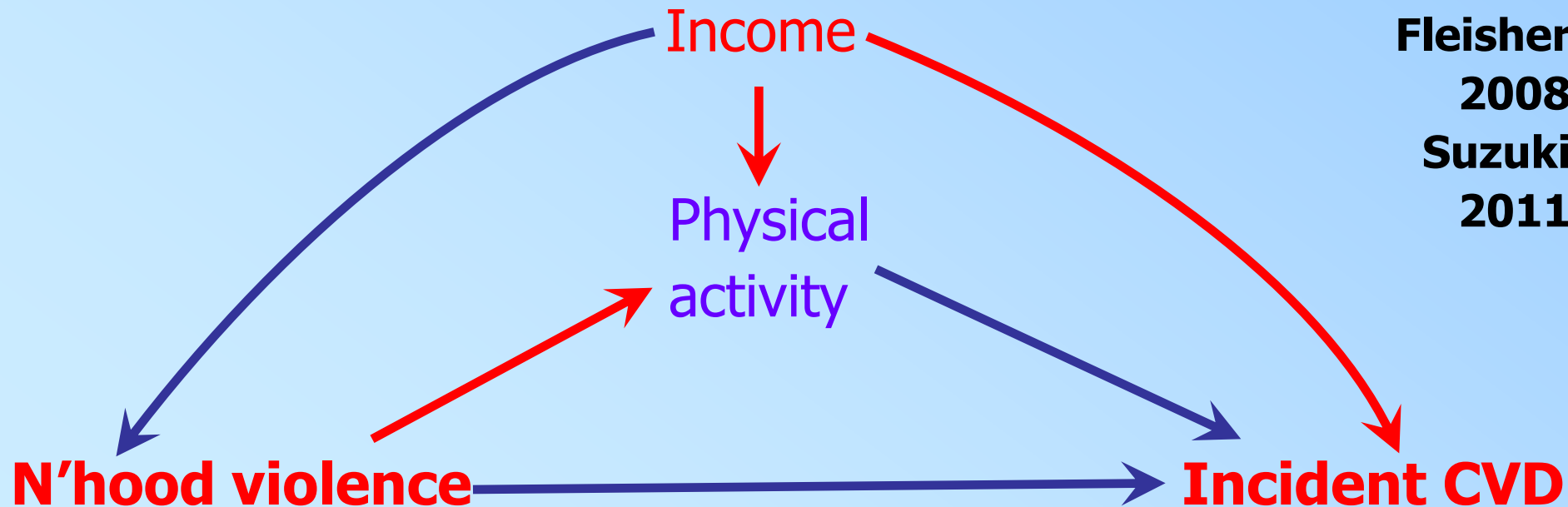
**Fleisher
2008
Suzuki
2011**



DAGs: TERMINOLOGY

- **A collider on a path:** a variable with two arrows into it (common effect): where two arrows “collide”
- **Unblocked path:** sequence of arrows connecting two variables that does not contain a collider

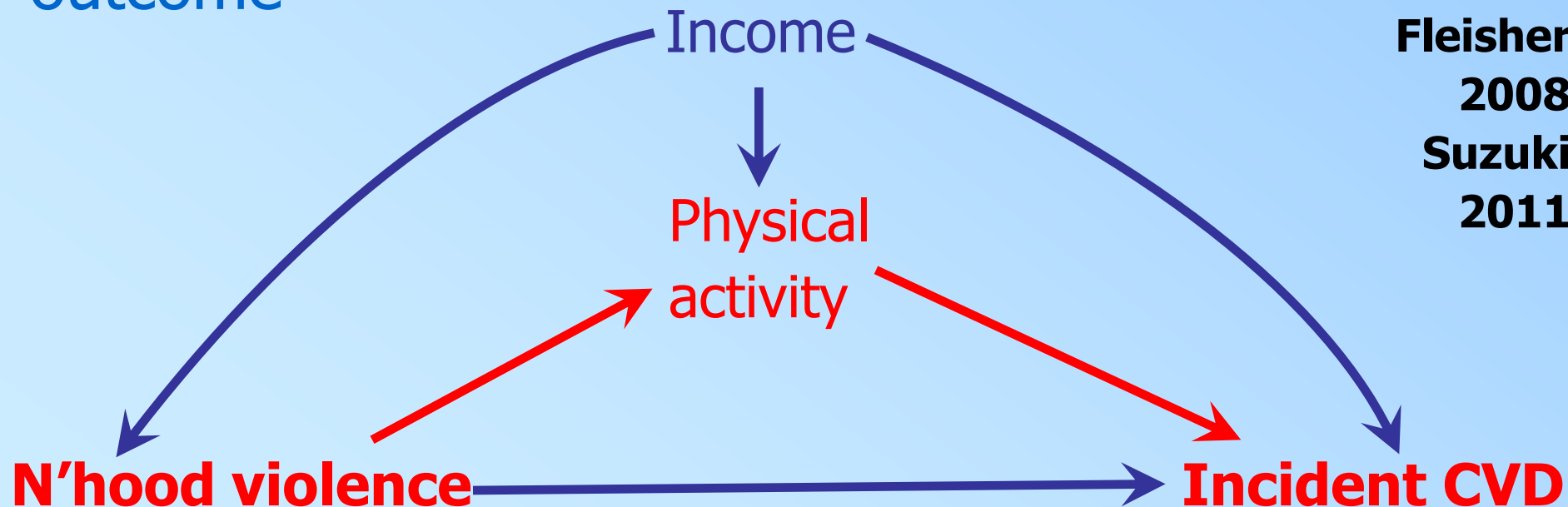
**Fleisher
2008
Suzuki
2011**



DAGs: TERMINOLOGY

- **Unblocked directed path**: sequence of directed arrows
- **Unblocked backdoor path**: an unblocked path that begins with an arrow pointing into the exposure and ends in an arrow pointing into the outcome

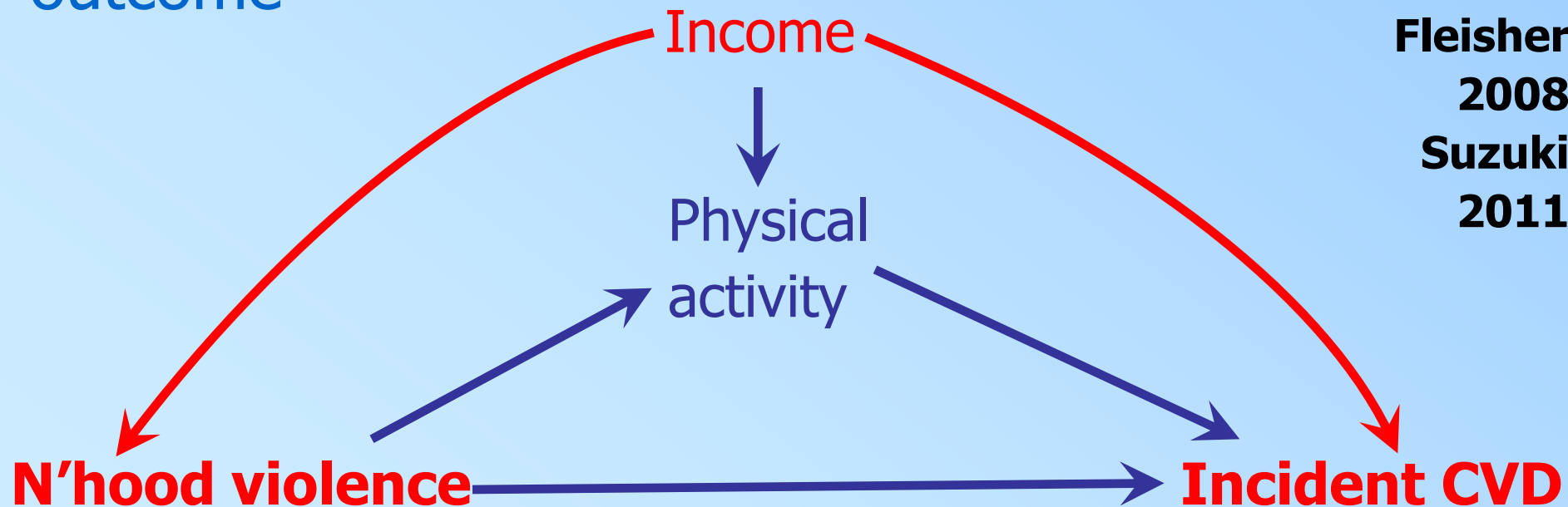
**Fleisher
2008
Suzuki
2011**



DAGs: TERMINOLOGY

- **Unblocked directed path**: sequence of directed arrows
- **Unblocked backdoor path**: an unblocked path that begins with an arrow pointing into the exposure and ends in an arrow pointing into the outcome

**Fleisher
2008
Suzuki
2011**



D-SEPARATION RULES

Pearl 2000

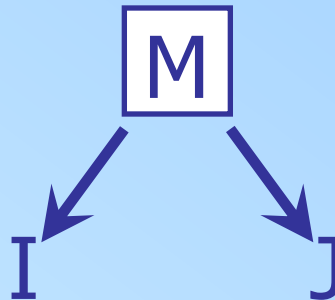
... or how to read the statistical independencies implied by the causal assumptions encoded in the DAG
= rules of dependence / independence of the nodes

Causal chain



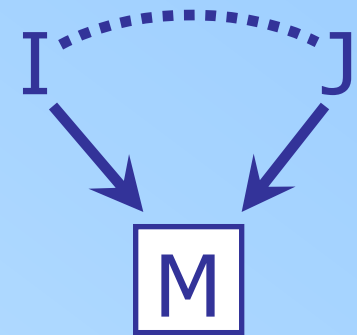
I and J independent
conditional on M
(d-separated):
directed path
blocked

Causal fork



I and J independent
conditional on M
(d-separated):
backdoor path
blocked

Inverted fork

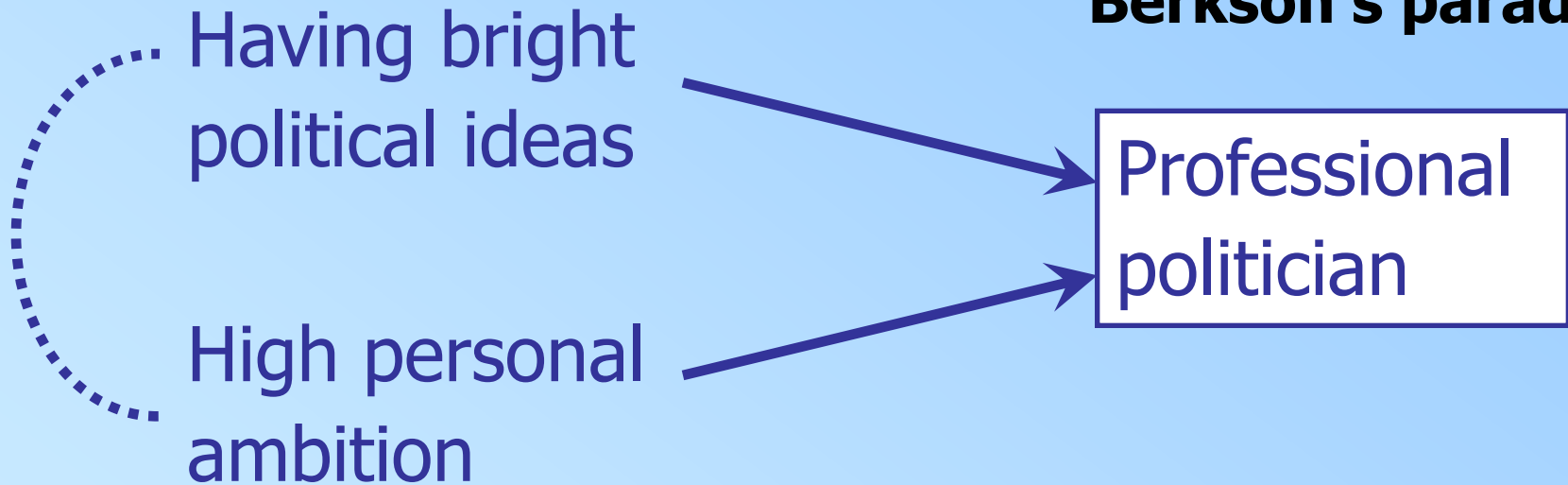


I and J dependent
conditional on M
(d-connected):
M is a collider

CONDITIONING ON A COLLIDER

Conditioning on a common effect of two variables induces an association between those variables.

“Berkson’s paradox”

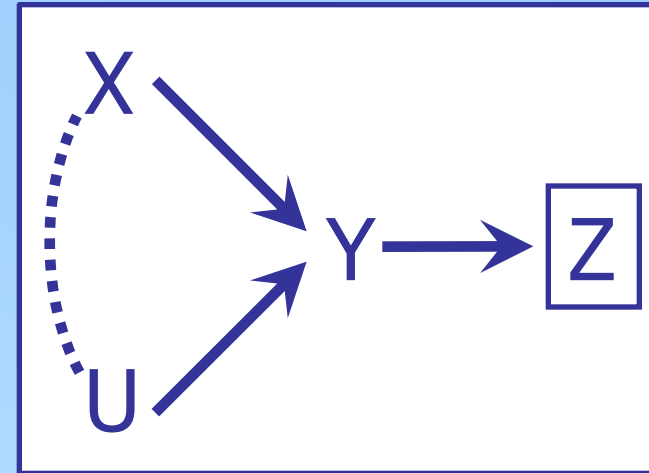


- Having bright political ideas and a high personal ambition are not associated in the whole population.
- Among professional politicians: if becoming a professional politician is not explained by one's bright ideas, then personal ambition is likely to be present...

D-SEPARATION RULES

The d-separation rules imply the following:

- XY marginally dependent
- XZ marginally dependent
- XZ independent conditional on Y
- XU marginally independent
- XU dependent conditional on Y
- XU dependent conditional on Z



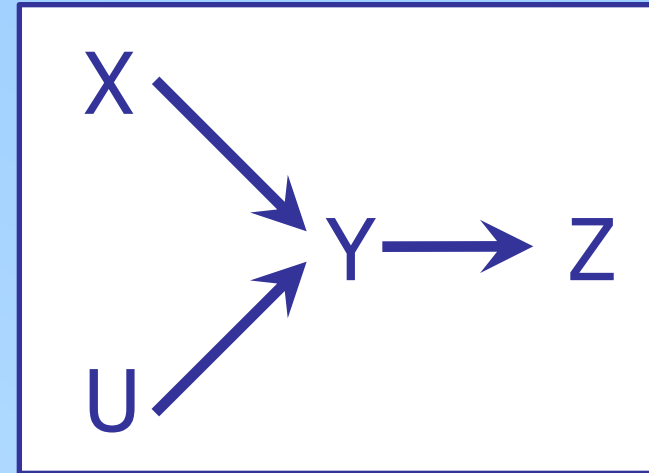
Pearl 2000
Glymour 2006

The Causal Markov Assumption: if we hold constant the direct causes of Y, any variable Z will be independent of Y, unless Z is an effect of Y

D-SEPARATION RULES

The d-separation rules imply the following:

- XY marginally dependent
- XZ marginally dependent
- XZ independent conditional on Y
- XU marginally independent
- XU dependent conditional on Y
- XU dependent conditional on Z



Pearl 2000
Glymour 2006

The Causal Markov Assumption: if we hold constant the direct causes of Y, any variable Z will be independent of Y, unless Z is an effect of Y

CLASSIFICATION OF BIAS

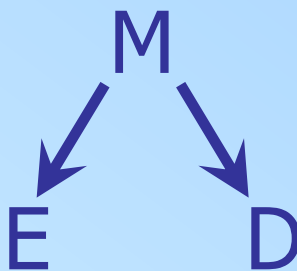
Apart from random variations, 3 basic causal structures (and more complex ones) can explain an association between an exposure (E) and a disease (D):

Cause and
effect

$E \longrightarrow D$

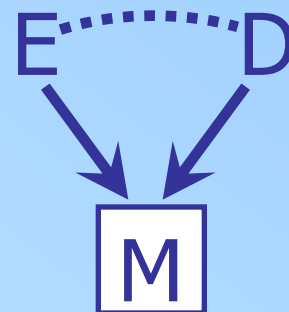
$D \longrightarrow E$

Common
cause



Confounding bias:
there is a common
cause of E and D

Common
effect



**Selection bias
(collider bias):**
conditional association
within strata of a
common effect

**Hernán. A
structural approach
to selection bias.
Epidemiology 2004.**

APPLICATION of DAGs in N'HOOD & HEALTH RESEARCH

1. Identifying variables that need to be adjusted for in estimating n'hood health effects
2. Why adjusting for a mediator does not necessarily estimate the direct n'hood effect?
3. Why sample selection results in spurious associations between n'hoods and health?

Glymour 2006
Fleisher 2008

ADJUSTMENT OF N'HOOD EFFECTS (1)

Aim: identify the set S of variables that needs to be adjusted for to estimate the (total) causal effect

“Backdoor test for sufficiency”:

S is sufficient for adjustment...

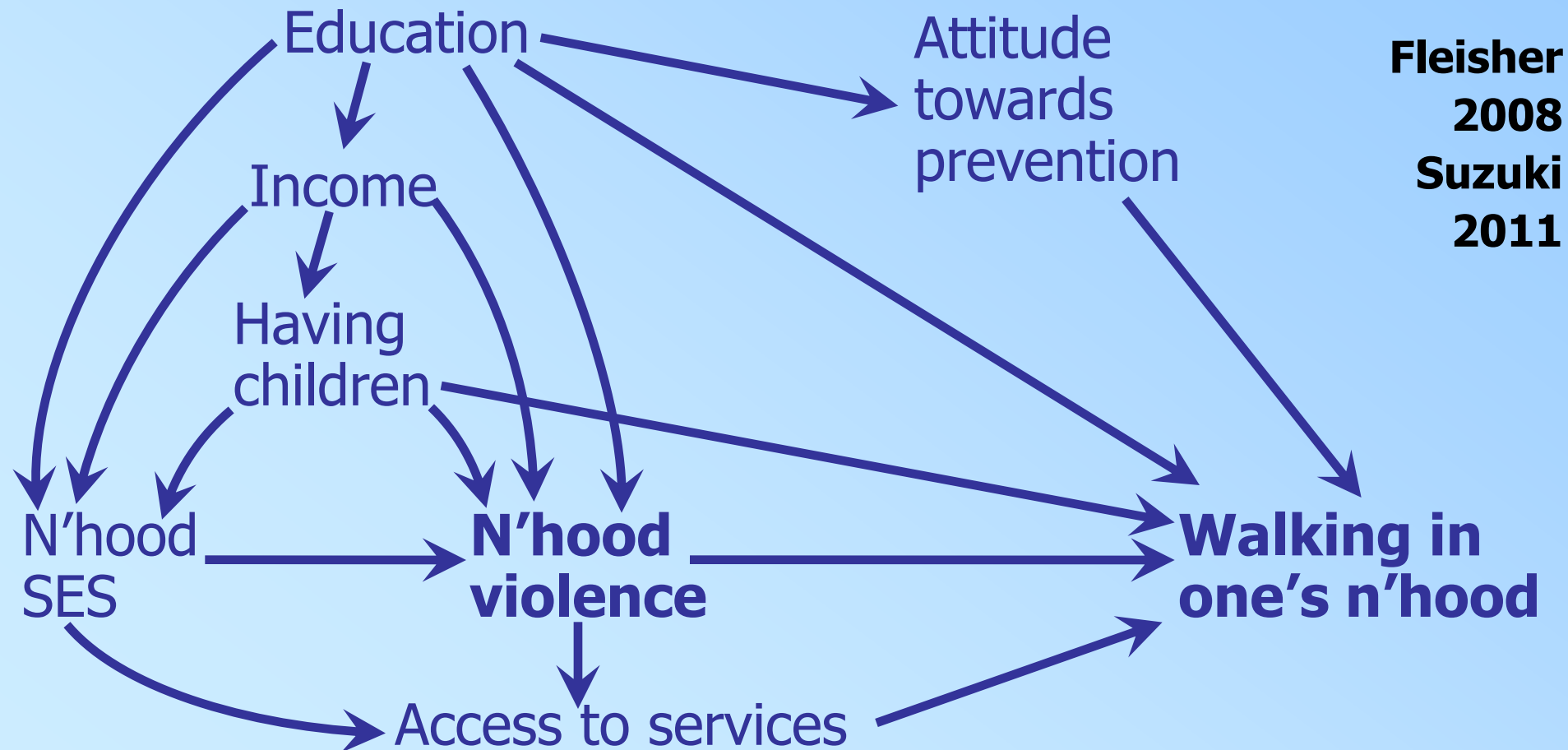
- if no variable in S is a descendant of the exposure (to avoid overadjustment) or the outcome
- if every unblocked backdoor path between the exposure and the outcome is intercepted by a variable in S

If there is a collider on a exposure-outcome path:

- we must not condition either on the collider or on any of its descendants
- **or** every unblocked path induced by adjustment for the collider must be intercepted by a variable in S

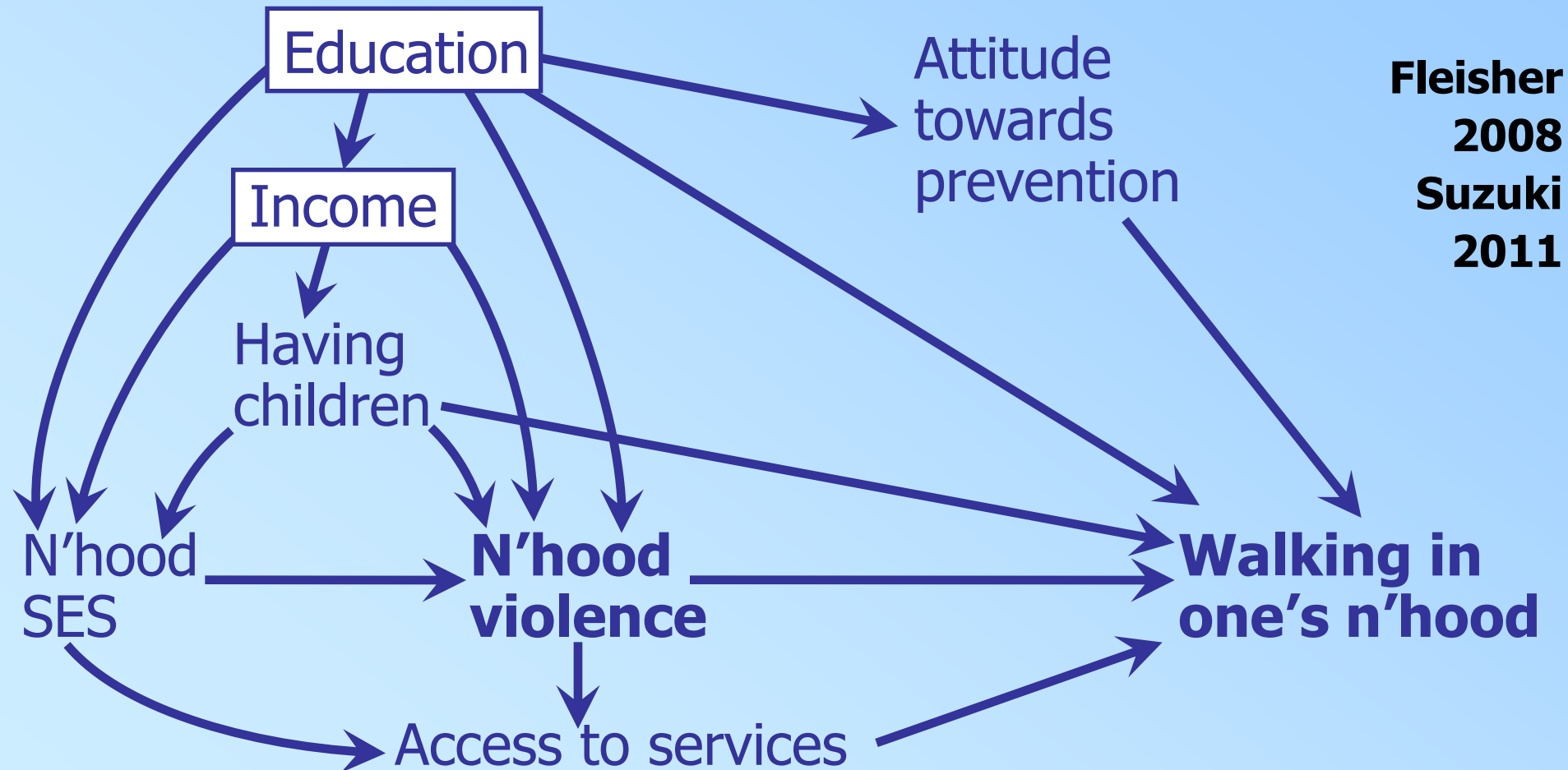
ADJUSTMENT OF N'HOOD EFFECTS (2)

Steps to determine the set of covariates S:



ADJUSTMENT OF N'HOOD EFFECTS (2)

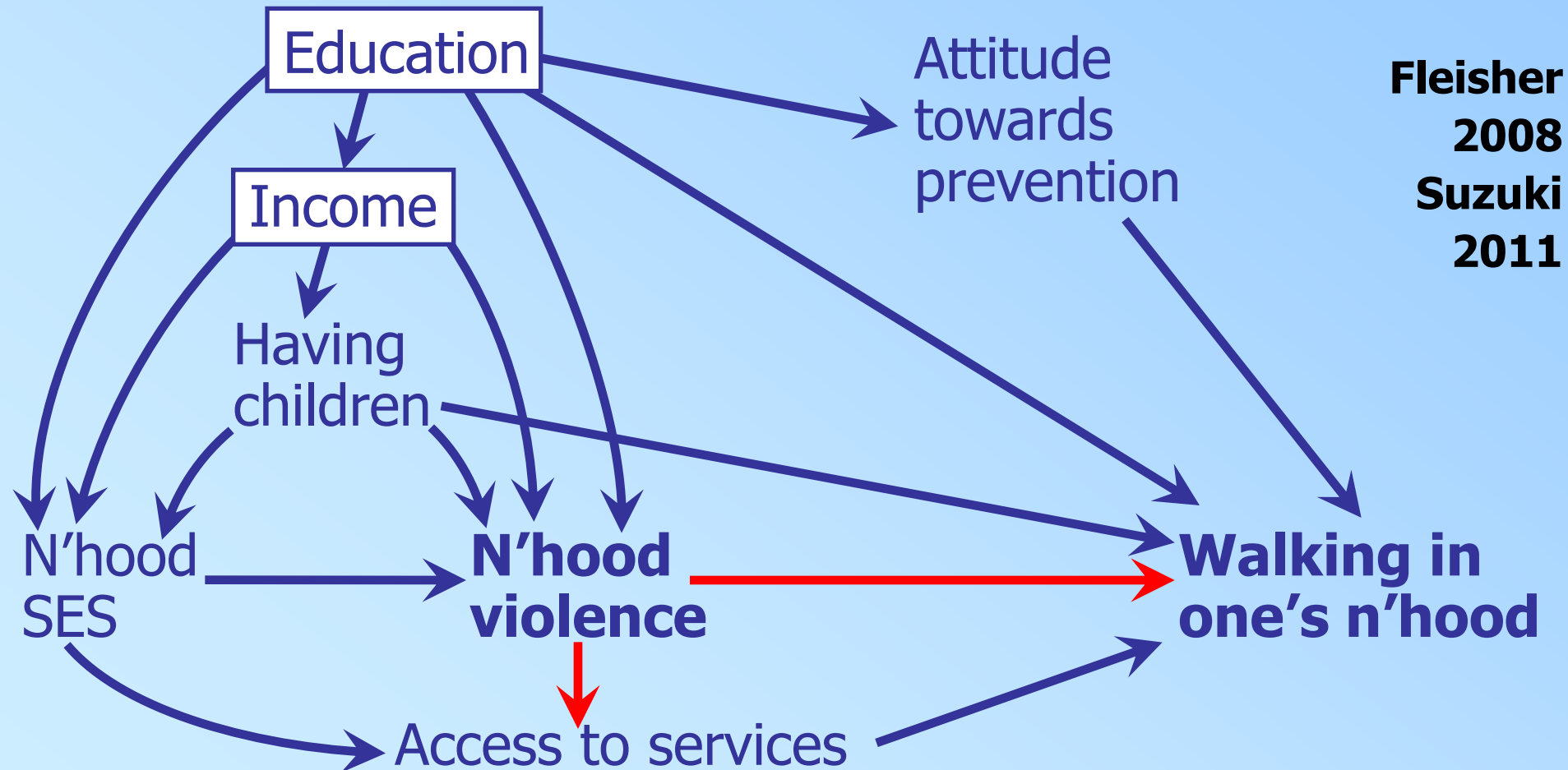
Steps to determine the set of covariates S:



ADJUSTMENT OF N'HOOD EFFECTS (2)

Steps to determine the set of covariates S:

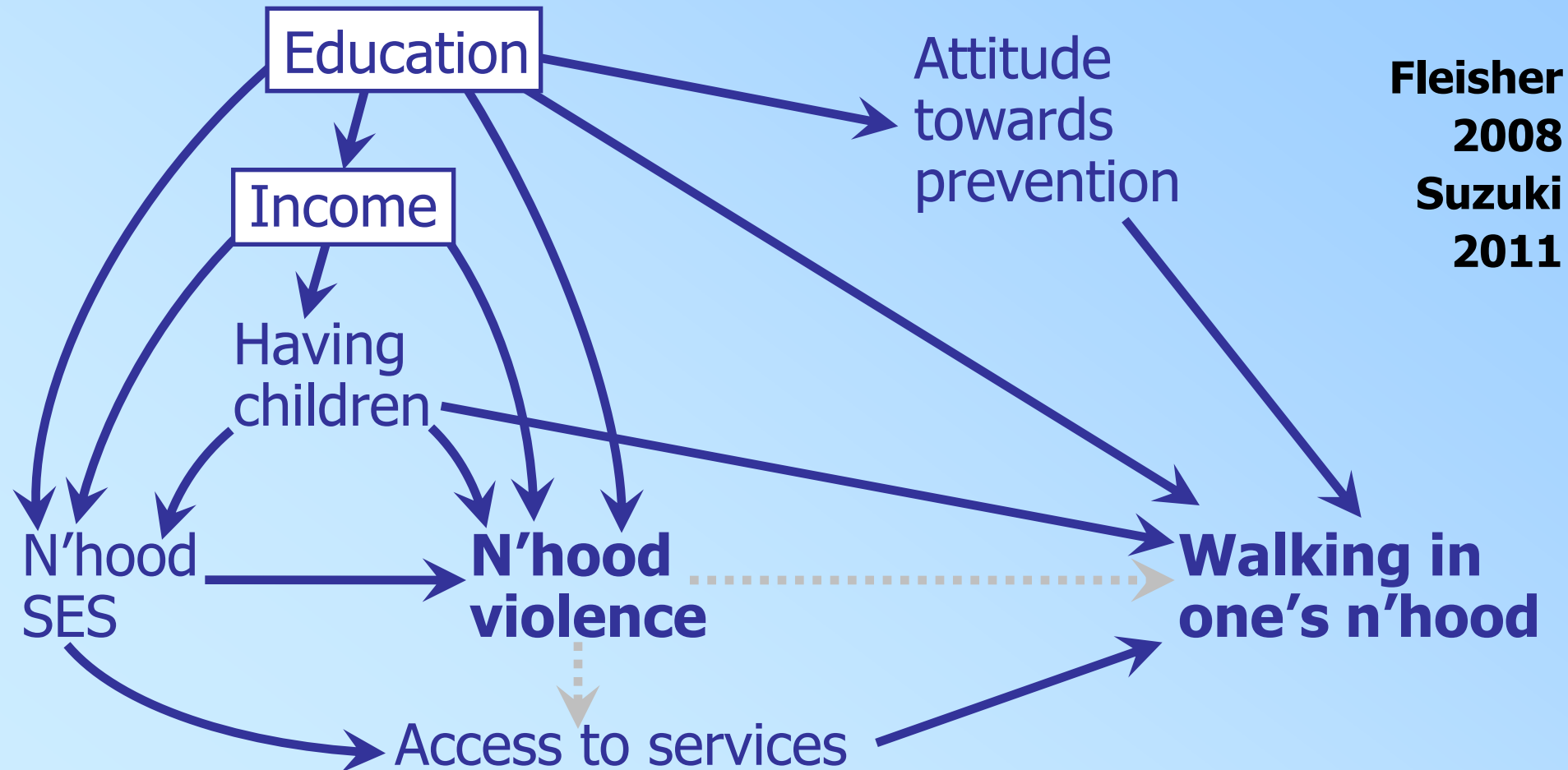
- 1) Delete all arrows emanating from the exposure



ADJUSTMENT OF N'HOOD EFFECTS (2)

Steps to determine the set of covariates S:

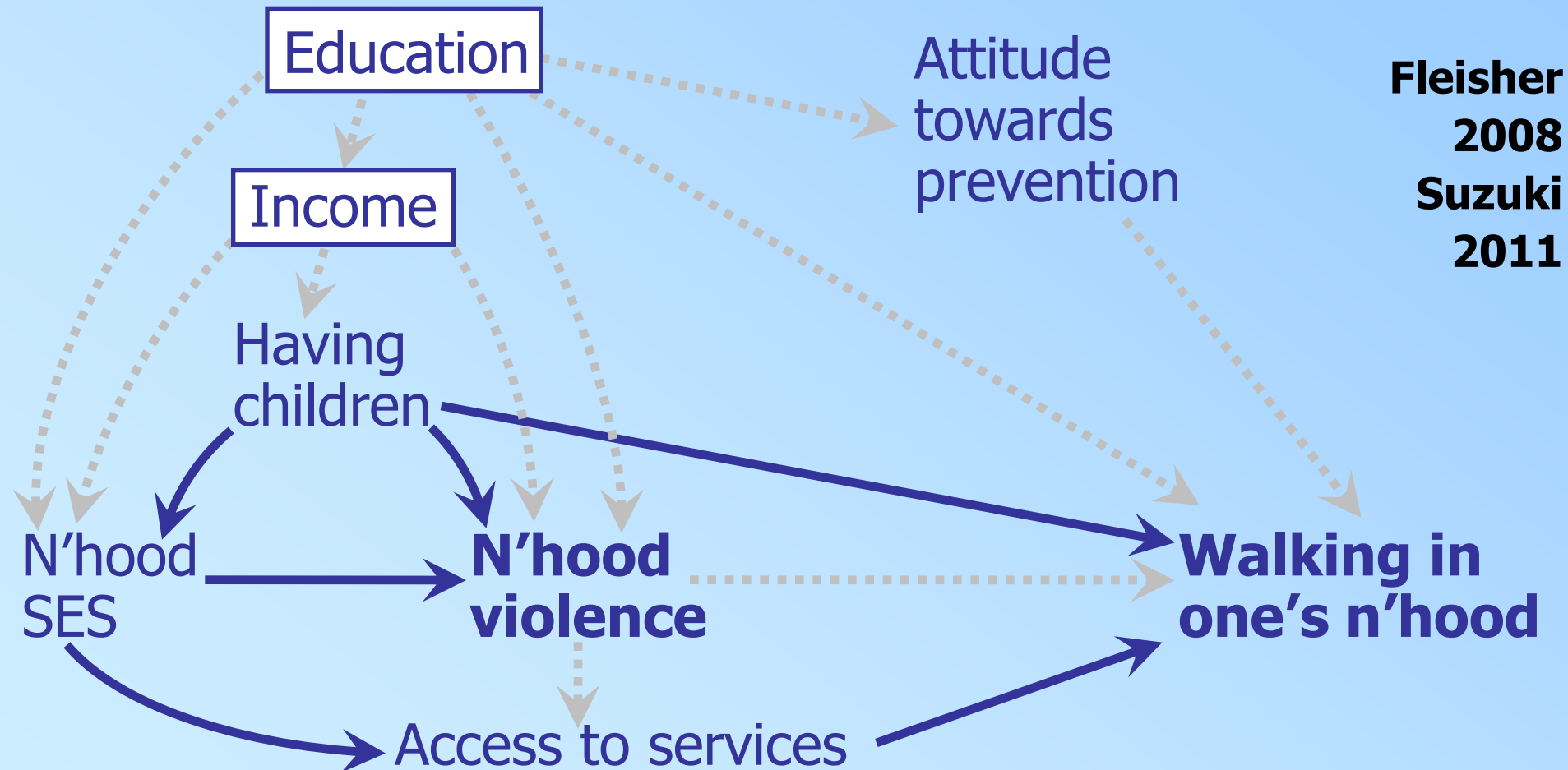
2) Draw undirected arcs to connect every variable that share a child or a descendent in S



ADJUSTMENT OF N'HOOD EFFECTS (2)

Steps to determine the set of covariates S:

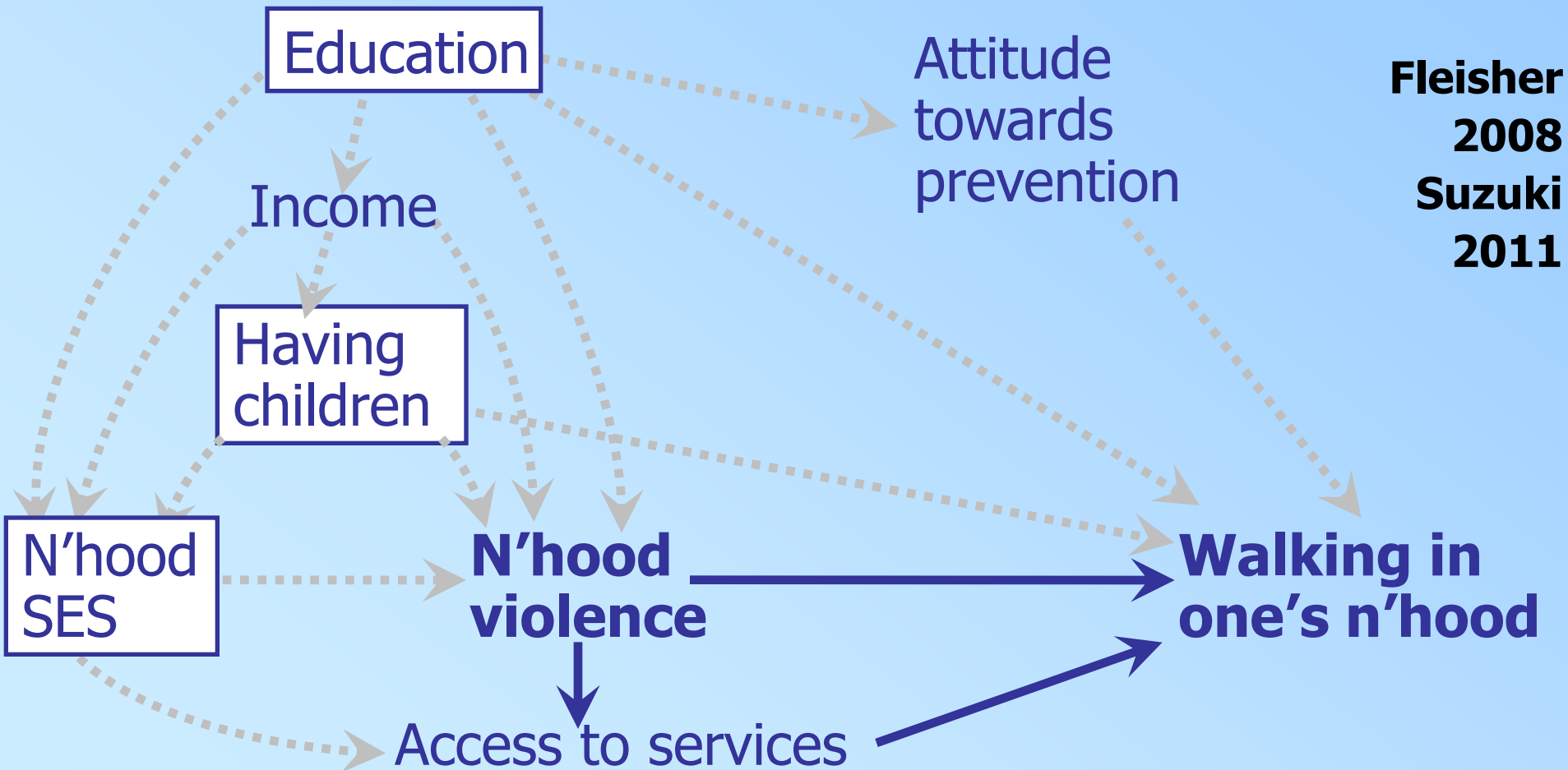
3) Mentally remove backdoor paths that are blocked by a variable in S



ADJUSTMENT OF N'HOOD EFFECTS (2)

Steps to determine the set of covariates S:

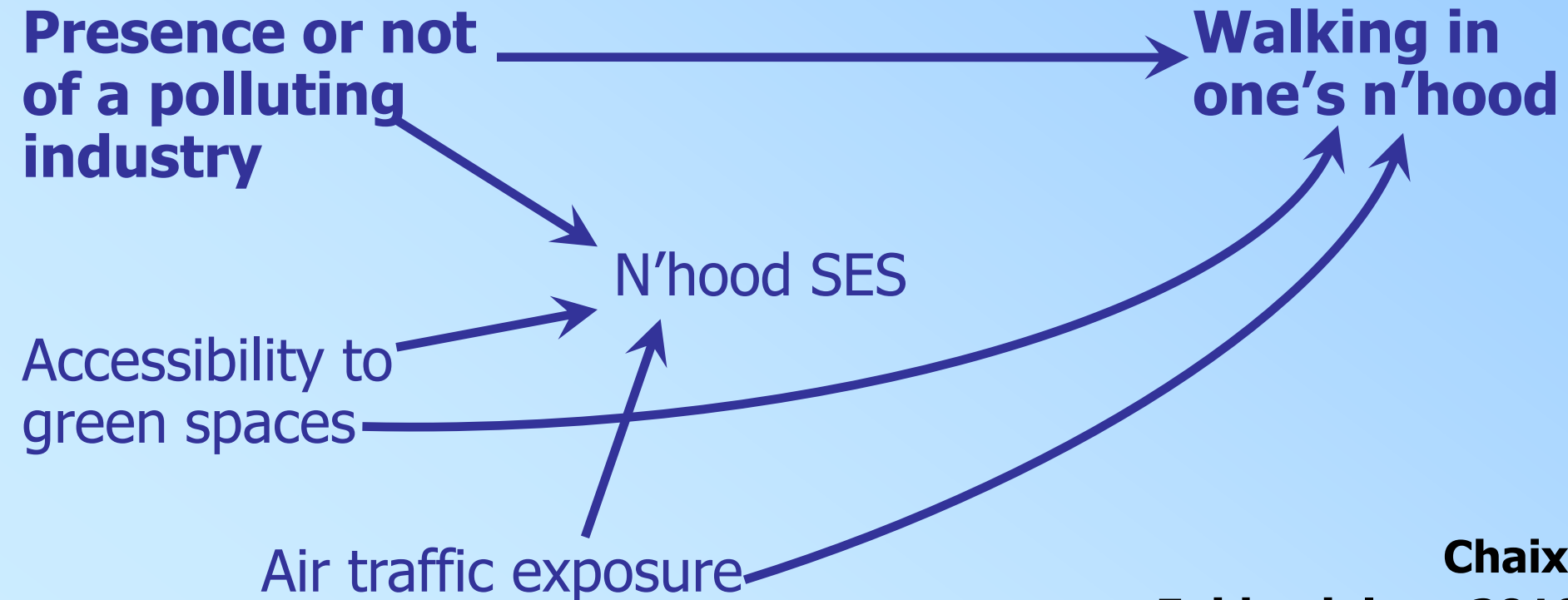
4) Redefine the adjustment set and identify the “minimally sufficient adjustment set”



ADJUSTMENT OF N'HOOD EFFECTS (3)

Steps to determine the set of covariates S:

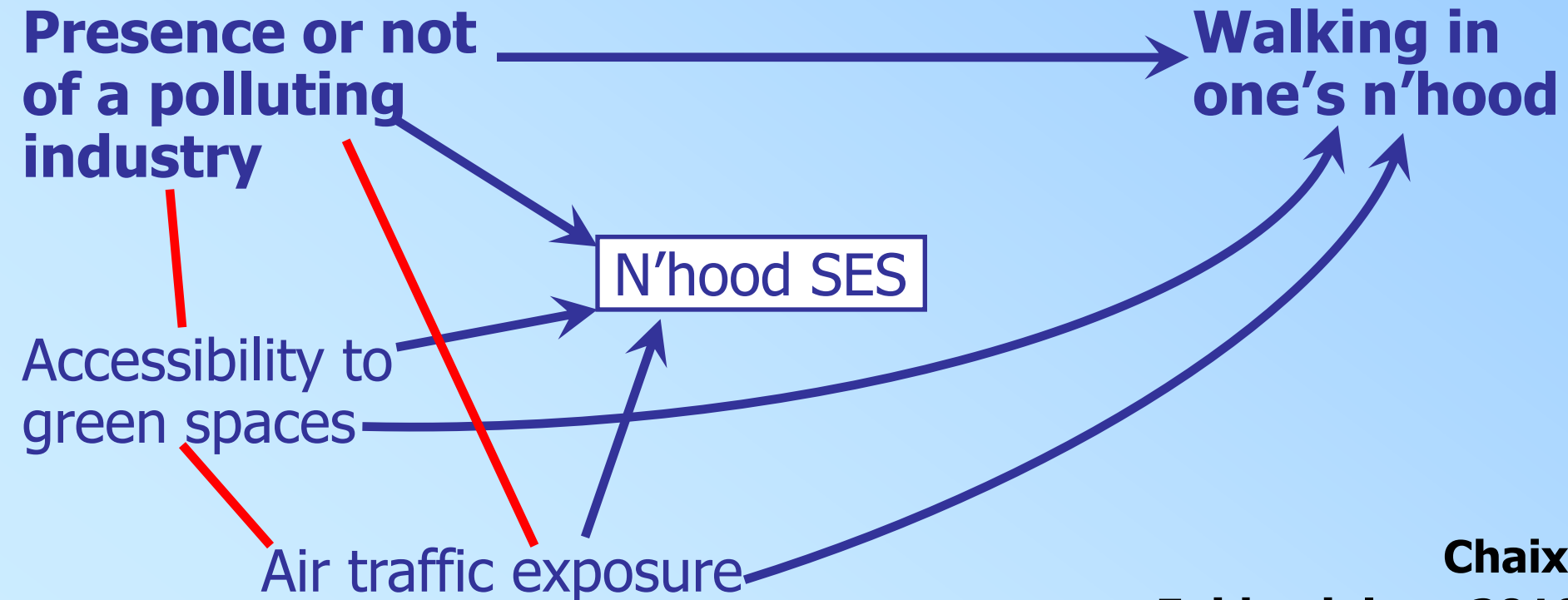
2) Draw undirected arcs to connect every variable that share a child or a descendent in S



ADJUSTMENT OF N'HOOD EFFECTS (3)

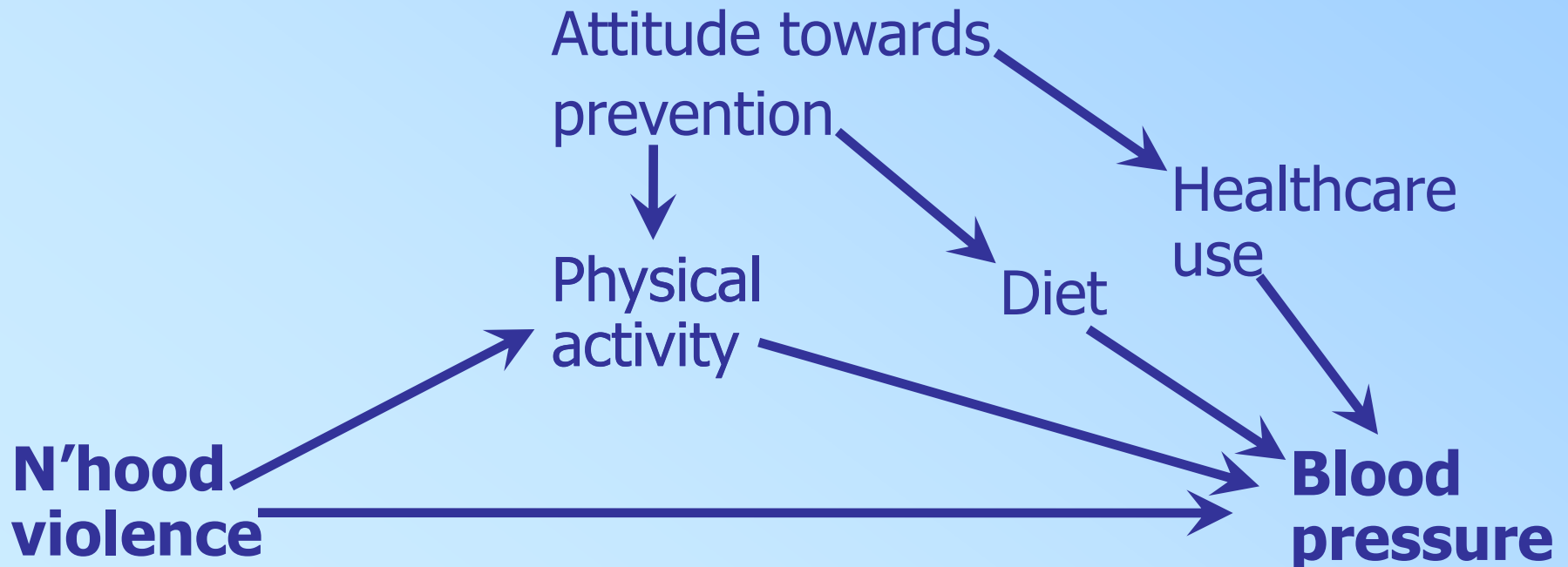
Steps to determine the set of covariates S:

2) Draw undirected arcs to connect every variable that share a child or a descendent in S



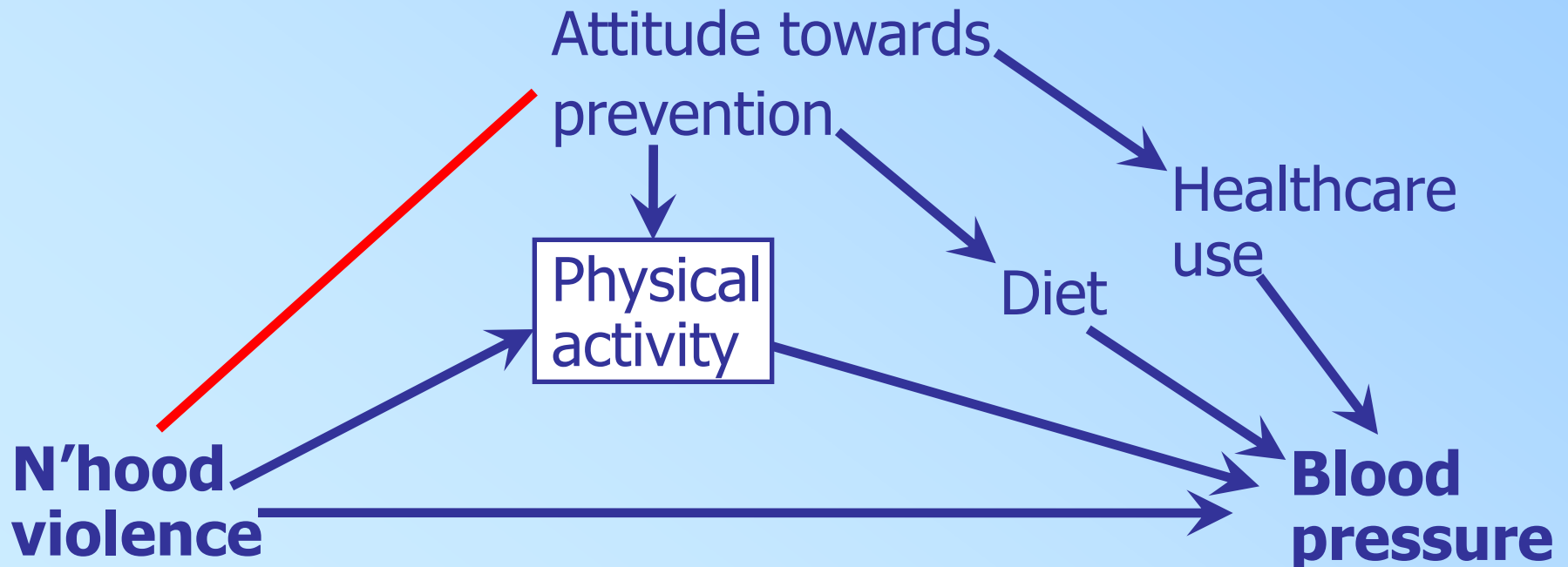
Estimating a direct effect by conditioning on a mediator

- The total effect is not confounded.
- In case of confounding between the mediator and the outcome, adjusting for the mediator (as a collider) will induce a spurious association between the exposure and outcome.



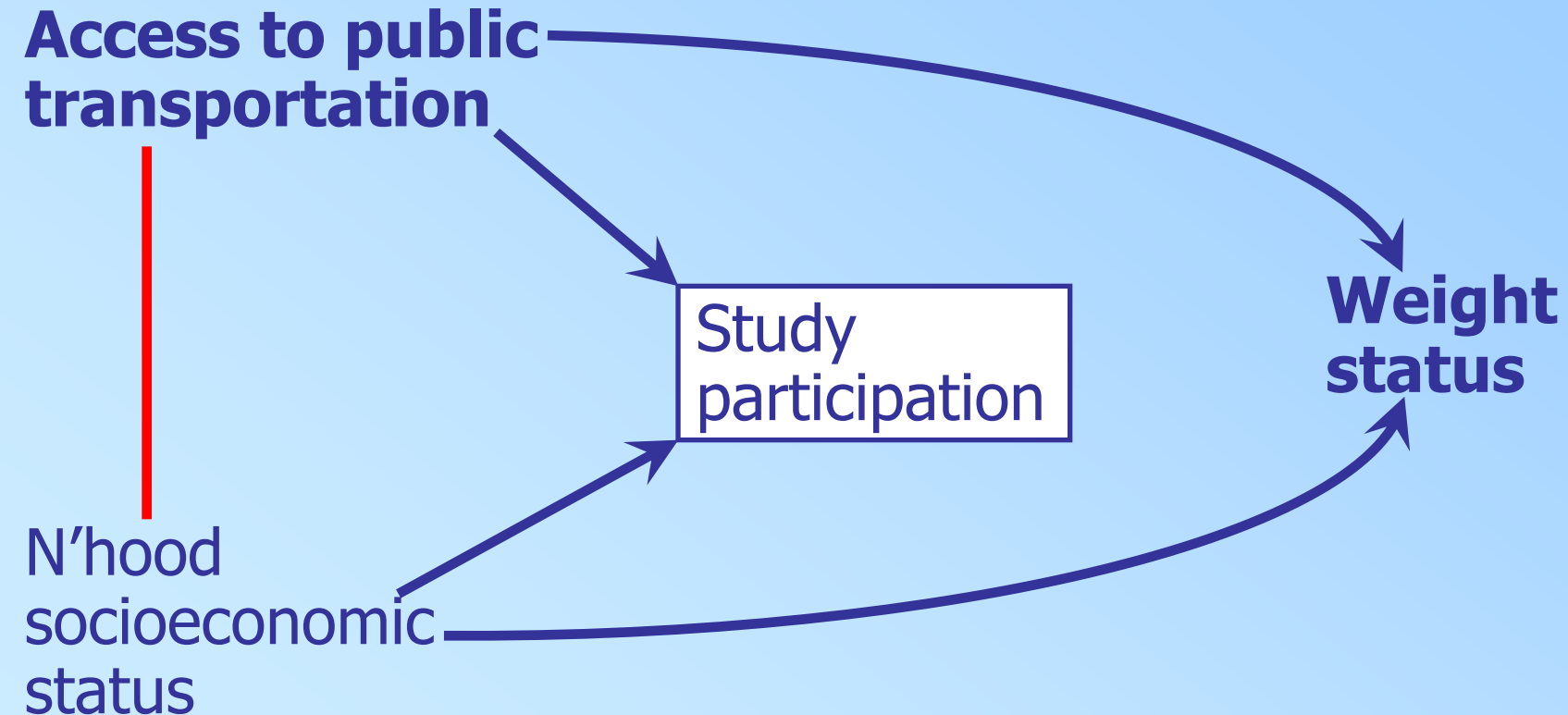
Estimating a direct effect by conditioning on a mediator

- The total effect is not confounded.
- In case of confounding between the mediator and the outcome, adjusting for the mediator (as a collider) will induce a spurious association between the exposure and outcome.



SAMPLE SELECTION BIAS

Sample selection bias may occur if study participation depends on both the exposure and the outcome or their causes.

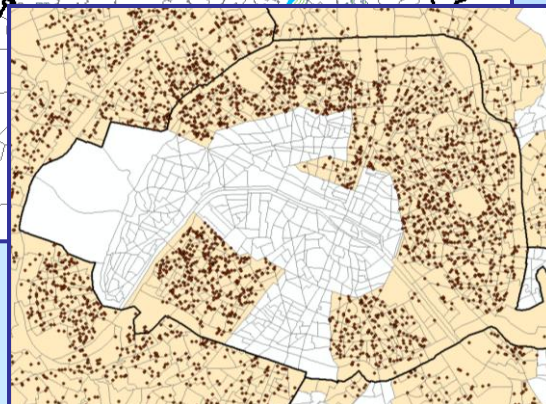
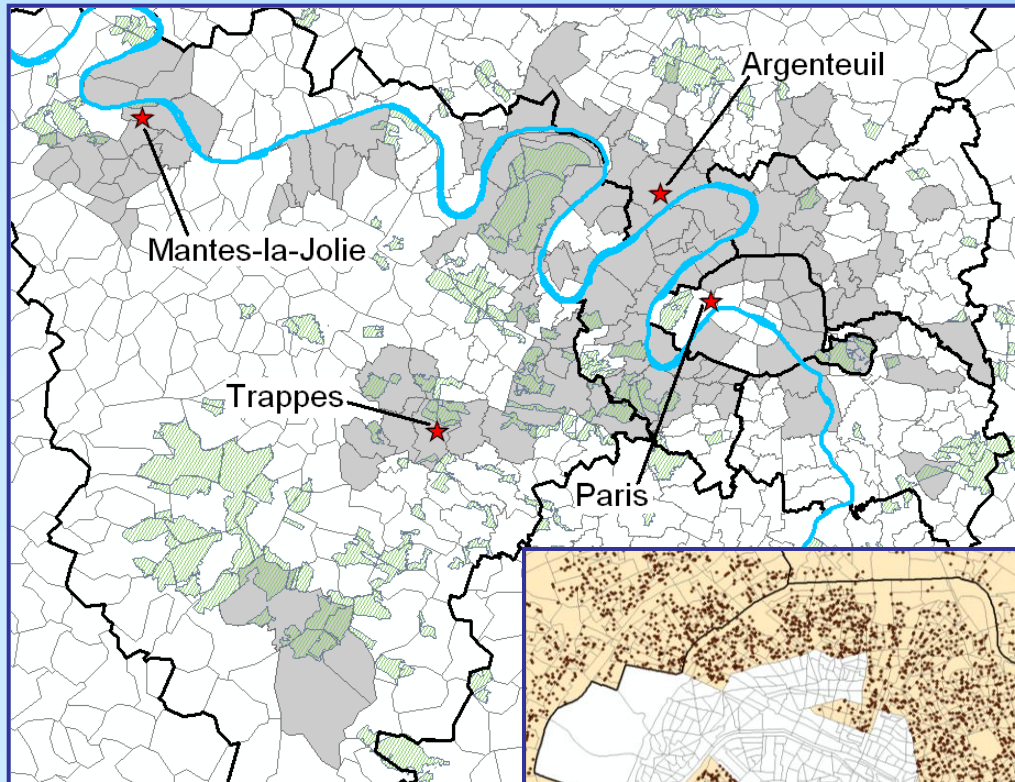


APPLICATION: THE RECORD STUDY

Recruitment during health checkups

- **7290** participants (30–79 years)

111 municipalities + 10 districts of Paris
= 1915 different neighborhoods



RECORD Study, wave 1

Biological data

Paramedical examinations

Medical questionnaires

Address & contact info 

RECORD questionnaire

Geocoding of participants

Environmental data

Healthcare use (SNIR-AM)

Hospitalizations (PMSI)

Mortality (Insee, CepiDC)

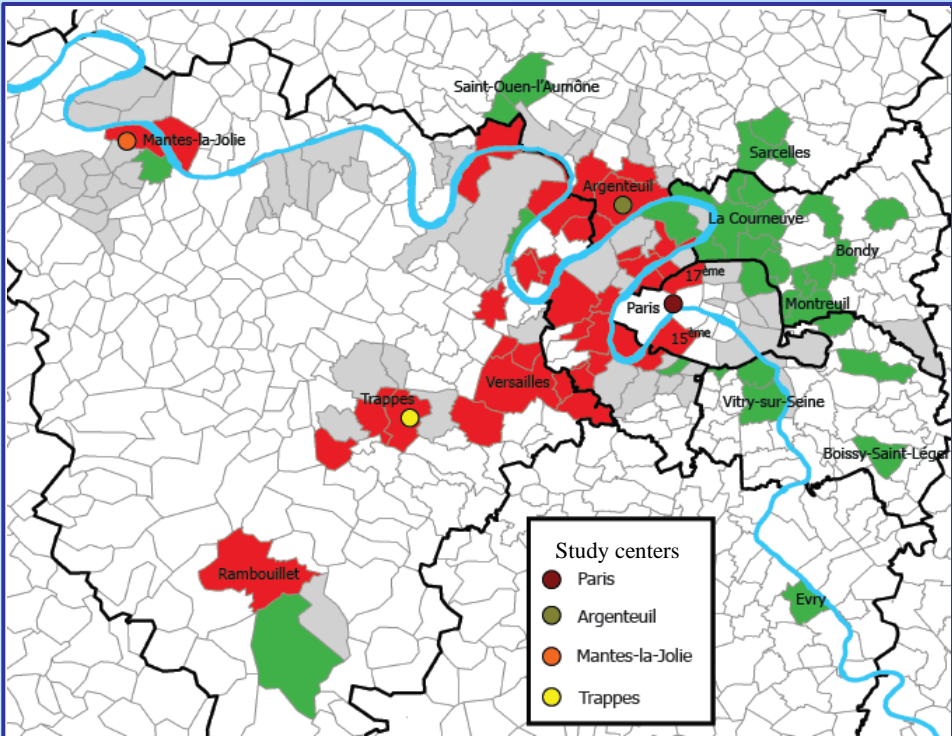
Professional career (CNAV)



MODELING OF STUDY PARTICIPATION

Multilevel Poisson model for participation of populations in the RECORD Cohort Study

Relatively large variance between neighborhoods



Rate of participation in the RECORD Study estimated from age/sex adjusted multilevel models

■ High participation (+25% and more) ■ Intermediary participation
■ Low participation (-25% and more) □ Out of study territory

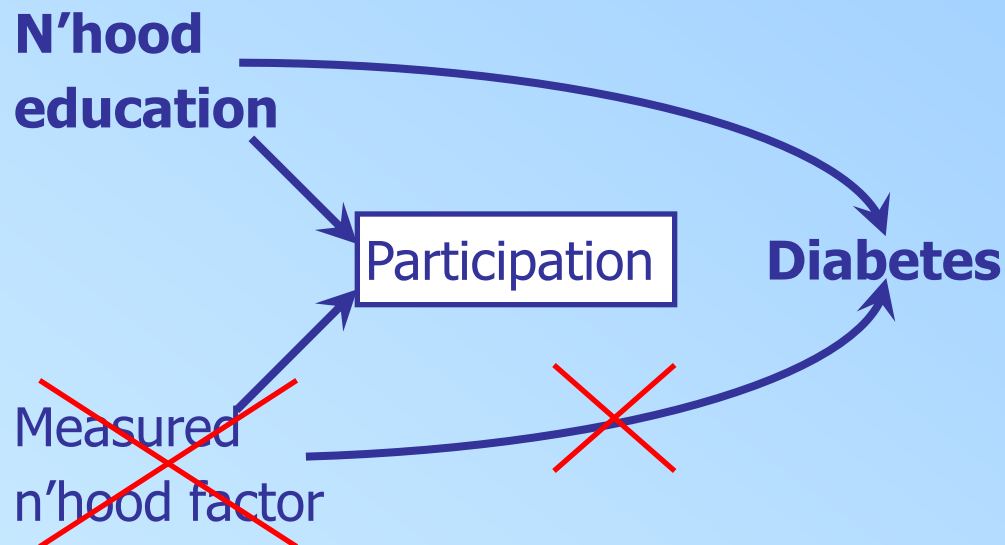
	PRR* (95% CI)
Individual education (vs. low)	
Medium	1.90 (1.74, 2.08)
High	4.25 (3.87, 4.67)
Distance to the center (vs. long)	
Medium-long	1.19 (1.09, 1.30)
Medium-short	1.45 (1.32, 1.58)
Short	1.75 (1.60, 1.91)
Median income (vs. low)	
Medium-low	1.20 (1.09, 1.32)
Medium-high	1.29 (1.14, 1.45)
High	1.39 (1.20, 1.60)
Mean real estate prices (vs. low)	
Medium-low	1.10 (1.00, 1.21)
Medium-high	1.11 (1.00, 1.24)
High	1.23 (1.09, 1.39)
% looking for work (vs. low)	
Medium-low	1.01 (0.93, 1.10)
Medium-high	1.18 (1.06, 1.31)
High	1.31 (1.15, 1.47)
% of area with buildings (vs. high)	
Medium-high	1.13 (1.03, 1.23)
Medium-low	1.26 (1.14, 1.39)
Low	1.37 (1.23, 1.51)
Building height (vs. high)	
Medium-high	1.11 (1.03, 1.21)
Medium-low	1.27 (1.16, 1.39)
Low	1.27 (1.15, 1.40)

Neighborhood education and type 2 diabetes

- Weak association between neighborhood average education and prevalence of type 2 diabetes after adjustment for individual socioeconomic characteristics

Association between n'hood education and diabetes		
Neighborhood education (vs. high)	OR	(95% CrI)
Medium-high	1.05	(0.70 – 1.56)
Medium-low	1.19	(0.80 – 1.75)
Low	1.56	(1.06 – 2.31)

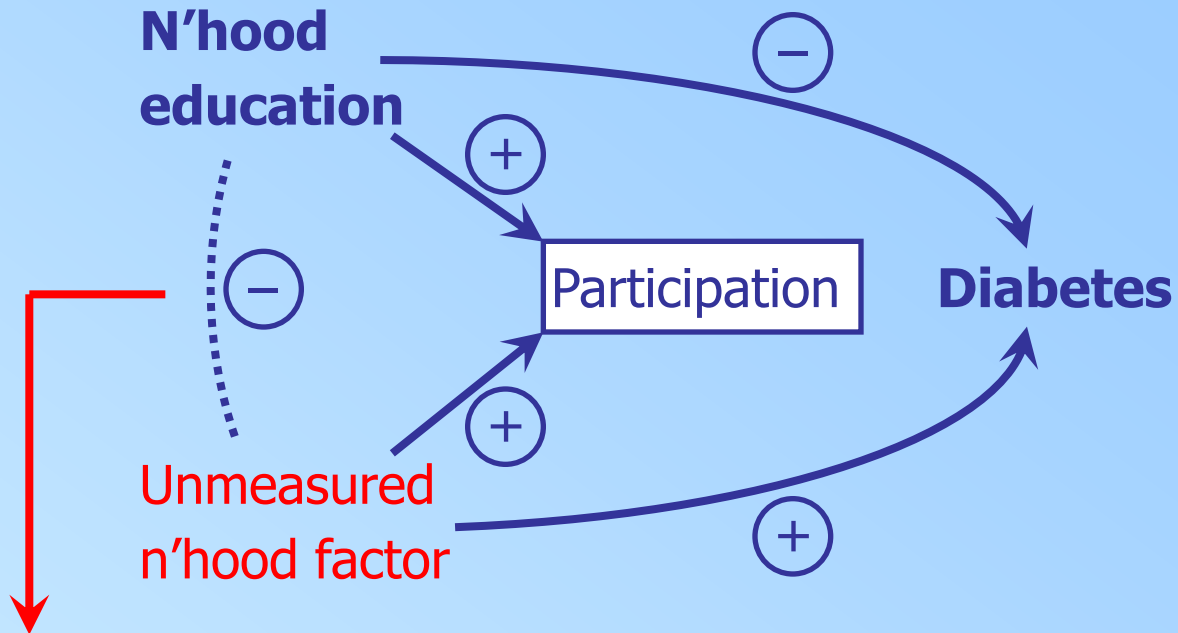
- None of the identified n'hood determinants of study participation biased the relationship of interest



Neighborhood education and type 2 diabetes

- Unmeasured n'hood determinants of participation could also bias the relationship between n'hood education and diabetes: **positive association with diabetes**

- Unmeasured n'hood determinants of participation were assessed with the **n'hood random effect** of the model for participation



- Correlation between n'hood education and the participation random effect:

- in the population: $r = -0.004$ ($-0.005, -0.002$) [$N = 3.1$ m]
- in the sample: $r = -0.14$ ($-0.17, -0.12$) [$N = 7233$]

Neighborhood education and type 2 diabetes

→ Adjust for the random effect reflecting variations in participation (a model-based value implying uncertainty)

Modeling inspired from: **Heckman JJ. Sample selection bias as a specification error. *Econometrica* 1979;47:153-61.**

Joint estimation of the 2 models through MCMC:

Model for participation

$$\text{Log}(\lambda_{ij}) = \beta_0 + \sum \beta_i X_i + s_j$$

Model for diabetes

$$\text{Logit}(p_{ij}) = \beta'_0 + \sum \beta'_i X_i + \gamma s_j + u_j$$

	Initial model		Model with correction	
Neighborhood education (vs. high)	OR	(95% CrI)	OR	(95% CrI)
Medium-high	1.05	(0.70 – 1.56)	1.01	(0.68 – 1.48)
Medium-low	1.19	(0.80 – 1.75)	1.15	(0.78 – 1.69)
Low	1.56	(1.06 – 2.31)	1.44	(0.98 – 2.13)

CONCLUSION

DAGs are useful because they challenge researchers:

- to formalize their research hypotheses
- to provide rationale for their analytic strategies
(ex: avoid the “kitchen sink approach” to adjustment)

Relevant developments based on DAGs:

- 1) Standardized methods to explore the coherence between alternative DAGs and empirical data (e.g., c-equivalence, **Pearl 2009**) (**D. Evans, EHESP**)
- 2) DAGs cannot encode assumptions on the strengths of associations → It is important to develop methods to place bounds on the amount of bias likely to be present under different assumptions.
- 3) Integration of interactions in DAGs, etc.

REFERENCES

- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37-48.
- Pearl J. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press, 2000.
- Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615-25.
- Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, eds. *Methods in Social Epidemiology* Jossey-Bass; 2006.
- Fleischer NL, Diez Roux AV. Using directed acyclic graphs to guide analyses of neighbourhood health effects: an introduction. *J Epidemiol Community Health* 2008;62:842-6.
- Pearl J, Paz A. Confounding Equivalence in Causal Inference. In: Grunwald P, Spirtes P, eds. *Proceedings of UAI*. Corvallis, OR: AUAI; 2010.
- Chaix B, Leal C, Evans D. Neighborhood-level Confounding in Epidemiologic Studies: Unavoidable Challenges, Uncertain Solutions. *Epidemiology* 2010;21:124-27.
- Chaix B, Billaudeau N, Thomas F, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology* 2011;22:18-26.