

Analyse de la médiation par g-computation : application dans la cohorte britannique NCDS 58

B. Lepage^{1,2,3}, D. Dedieu¹, V. Ehlinger¹, M. Kelly-Irving¹, N. Savy^{1,3,4}, T. Lang^{1,2,3}

1. Inserm UMR 1027, Toulouse
2. Service d'Epidémiologie, CHU de Toulouse
3. Faculté de médecine, Université Toulouse III Paul Sabatier
4. Institut de Mathématiques de Toulouse et Université de Toulouse

Contexte

- Dans la **cohorte de naissance britannique NCDS 58**
- **Mortalité augmentée en cas d'adversité pendant l'enfance (ACE)**
- ACE : score de 0 à 6 (6 dimensions)
« évènements ou situations intra familiaux entre 0 et 16 ans pouvant entraîner un stress aigu ou chronique chez l'individu ».

– Placement (foyer, famille d'accueil,...) ;

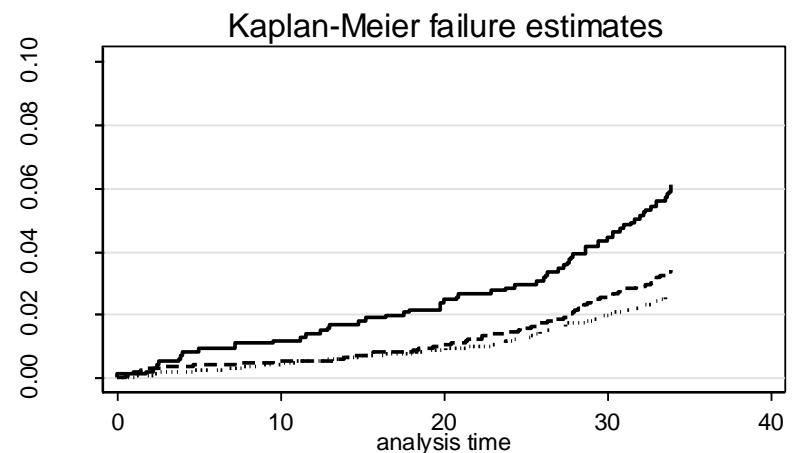
– Membre du foyer familial :

- en prison / probation
- séparation mère/père
- maladie mentale
- abus d'alcool

– Négligence physique

- **Après ajustement**

RR (ACE 2 vs 0) = 2,11 IC95% [1,59-2,80]



Number at risk

ace3cat	0	1	2
0	8573	2572	1019
1	8498	2547	1007
2	8407	2506	995

..... ACE=0 - - - - ACE=1 ——— ACE>1

Objectif

- Est-ce que cet effet de l'ACE sur la mortalité s'explique par des comportements intermédiaires entre 16 et 50 ans ?
- Médiateurs d'intérêt = Tabac, Alcool, Obésité
- Autre formulation : si on « intervenait » sur les médiateurs, de telle sorte que *tabac = alcool = obésité = NON* quel serait l'effet résiduel de l'ACE sur la mortalité ?
 - ⇒ Estimer l'effet direct contrôlé de l'ACE sur la mortalité pour une valeur fixée des médiateurs à *tabac < 1 cig/j ; alcool < 2-3 verres/j ; obésité = NON*

Estimation de l'effet direct contrôlé

- Classiquement, ajuster sur le médiateur ?

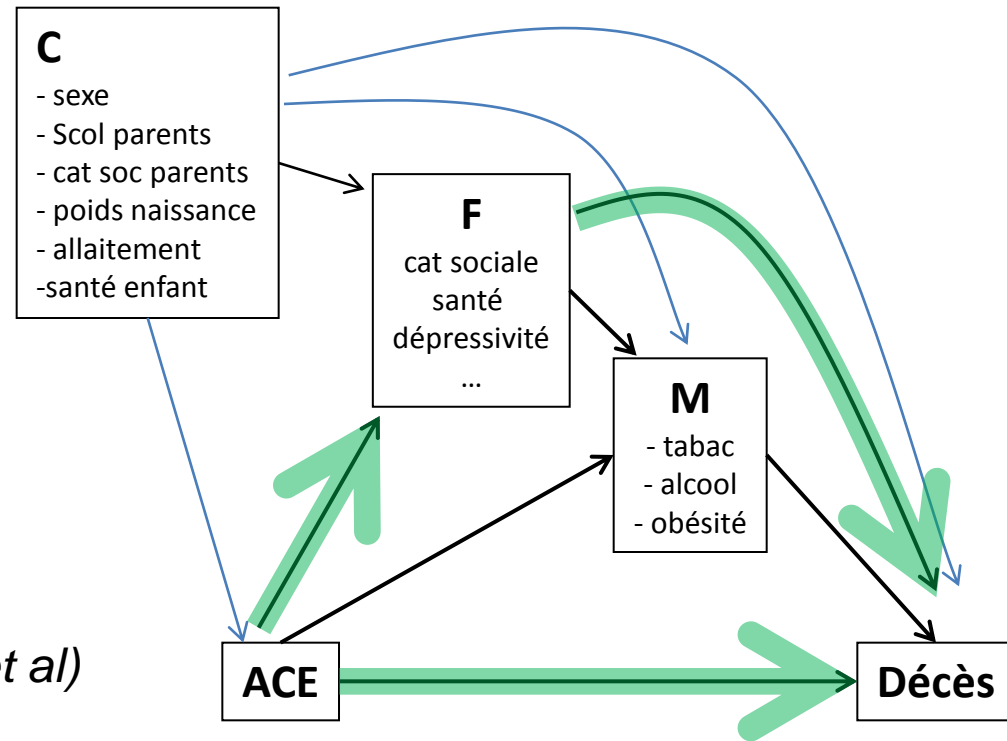
Par exemple avec un modèle de Cox :

$$h(t, ACE, M, C) = h_0(t) \exp(\mu_1 + \beta_{ACE} ACE + \beta_M M + \beta_C C)$$

méthode biaisée en cas de facteurs de confusion intermédiaire entre M et outcome

Méthodes alternatives pour estimer un effet direct contrôlé

- IPTW (modèles structurels marginaux) (*Robins, Vanderweele*)
- méthodes séquentielles (*Vansteelandt et al*)
- g-computation (*Robins*)



En vert : Effet direct contrôlé

Vague 3
16 ans

Vague 4
23 ans

Vague 5
33 ans

Vague 6
41/42 ans

Vague 7
46/47 ans

Vague 8
50 ans

0 à 16 ans

F 0
- sexe
- Scol parents
- cat soc parents
- poids naissance
- allaitement
-santé enfant

C3
Dépressivité
(16 ans)

- Scolarité max
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- handicap

ACE

M3
- tabac
- alcool
- obésité

M4
- tabac
- alcool
- obésité

M5
- tabac
- alcool
- obésité

M6
- tabac
- alcool
- obésité

M7
- tabac
- alcool

Décès
3-4

Décès
4-5

Décès
5-6

Décès
6-7

Décès
7-8

Vague 3
16 ans

Vague 4
23 ans

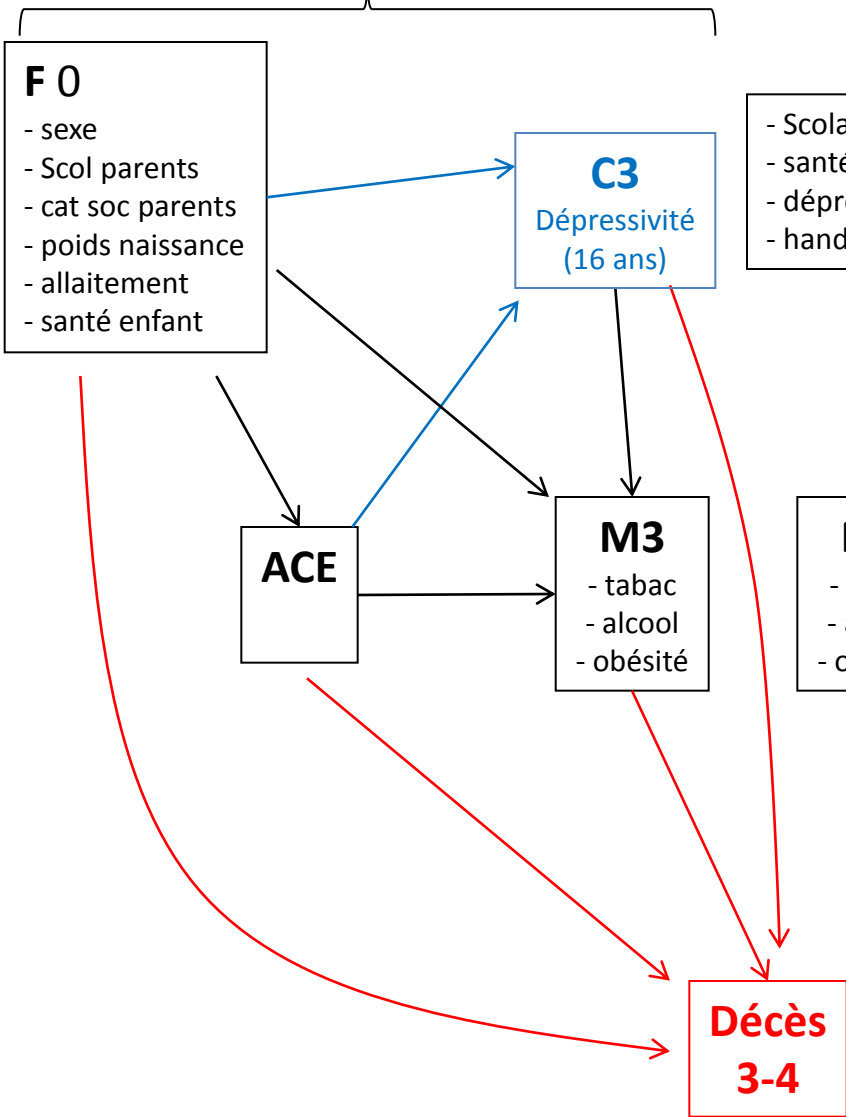
Vague 5
33 ans

Vague 6
41/42 ans

Vague 7
46/47 ans

Vague 8
50 ans

0 à 16 ans



Estimation de l'effet direct contrôlé : 2 étapes

1. Estimation des relations entre les variables

- facteurs de confusion intermédiaires en fonction des variables antérieures
 $P(\text{Dépressivité}_4 | F_0, ACE, C_3), \text{ etc}$
- décès en fonction des variables antérieures
 $P(\text{Décès}_4 | F_0, ACE, C_3, M_3), \text{ etc}$

2. Simulation de valeurs « contrefactuelles » des f. de confusion intermédiaires et des décès en fonction des valeurs fixées de l'ACE et des médiateurs, exemple :

- $dep_{doACE=0}$ à partir de la distribution $P(dep_4 | F_0, doACE=0)$
- $décès_{doACE=0, M3=0}$ à partir de la distribution $P(décès_4 | F_0, doACE=0, dep_{doACE=0}, doM_3=0)$

Vague 3
16 ans

Vague 4
23 ans

Vague 5
33 ans

Vague 6
41/42 ans

Vague 7
46/47 ans

Vague 8
50 ans

0 à 16 ans

F 0

- sexe
- Scol parents
- cat soc parents
- poids naissance
- allaitement
- santé enfant

Dépressivité (16 ans)

- Scolarité max
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- handicap

ACE

M3

- tabac
- alcool
- obésité

M4

- tabac
- alcool
- obésité

M5

- tabac
- alcool
- obésité

Et on continue d'une vague à l'autre

Les modèles se complexifient de plus en plus

Décès 4-5

Décès 5-6

Décès 6-7

Décès 7-8

Vague 3
16 ans

Vague 4
23 ans

Vague 5
33 ans

Vague 6
41/42 ans

Vague 7
46/47 ans

Vague 8
50 ans

0 à 16 ans

F 0,1,2,3
- sexe
- Scol parents
- cat soc parents
- poids naissance
- allaitement
- santé enfant

Dépressivité
(16 ans)

- Scolarité max
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- dépressivité
- handicap

- cat sociale
- santé perçue
- handicap

ACE

M3
- tabac
- alcool
- obésité

M4
- tabac
- alcool
- obésité

M5
- tabac
- alcool
- obésité

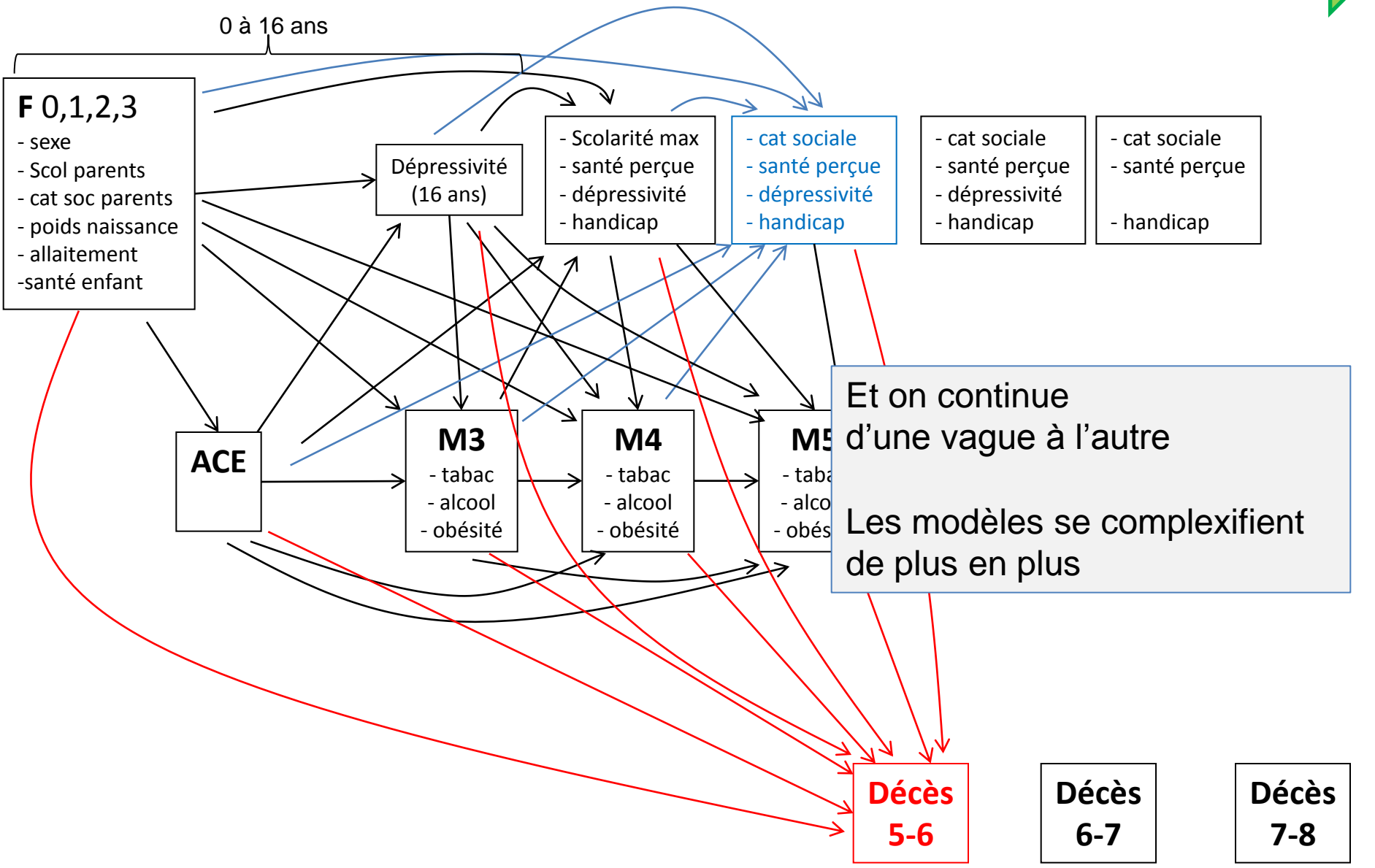
Et on continue
d'une vague à l'autre

Les modèles se complexifient
de plus en plus

**Décès
5-6**

Décès
6-7

Décès
7-8



Estimation de l'effet direct contrôlé

On peut donc simuler des « potential outcomes » ou « résultats contrefactuels » : **risque cumulés de décès avant 50 ans en fonction de différents niveaux d'interventions potentielles :**

$$R(\text{décès} \mid \text{doACE} = 0, \text{doTABAC}=0, \text{doALCOOL} = 0, \text{doOBESITE} = 0)$$

$$R(\text{décès} \mid \text{doACE} = 1, \text{doTABAC}=0, \text{doALCOOL} = 0, \text{doOBESITE} = 0)$$

$$R(\text{décès} \mid \text{doACE} = 2 \text{ ou } +, \text{doTABAC}=0, \text{doALCOOL} = 0, \text{doOBESITE} = 0)$$

Effet direct contrôlé si tous les médiateurs sont fixés à 0 :

	En différence de risque	En risque relatif
doACE=1 vs doACE=0	$\Delta = R(\text{décès} \mid \text{doACE} = 1, \text{doM} = 0) - R(\text{décès} \mid \text{doACE} = 0, \text{doM} = 0)$	$RR = \frac{R(\text{décès} \mid \text{doACE} = 1, \text{doM} = 0)}{R(\text{décès} \mid \text{doACE} = 0, \text{doM} = 0)}$
doACE=2 ou + vs doACE=0	$\Delta = R(\text{décès} \mid \text{doACE} = 2, \text{doM} = 0) - R(\text{décès} \mid \text{doACE} = 0, \text{doM} = 0)$	$RR = \frac{R(\text{décès} \mid \text{doACE} = 2, \text{doM} = 0)}{R(\text{décès} \mid \text{doACE} = 0, \text{doM} = 0)}$

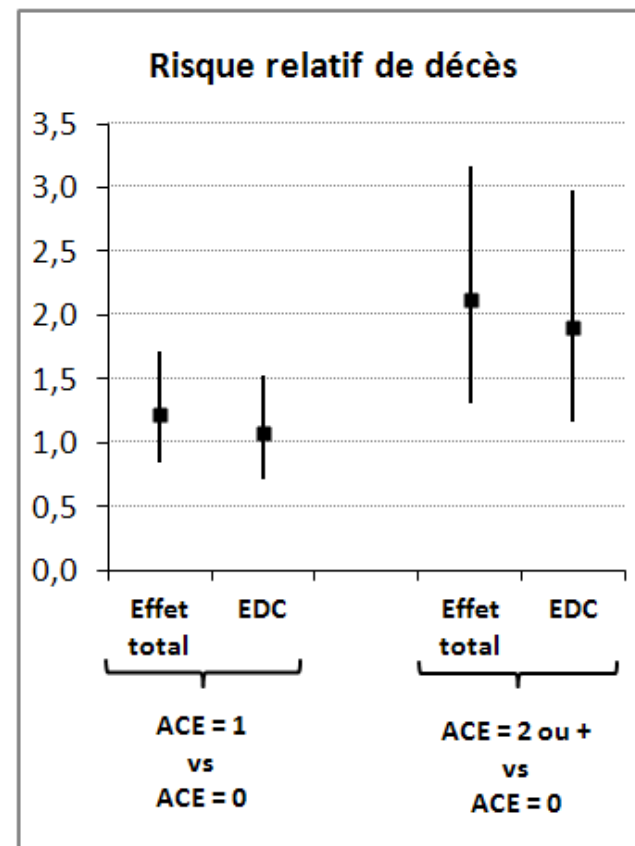
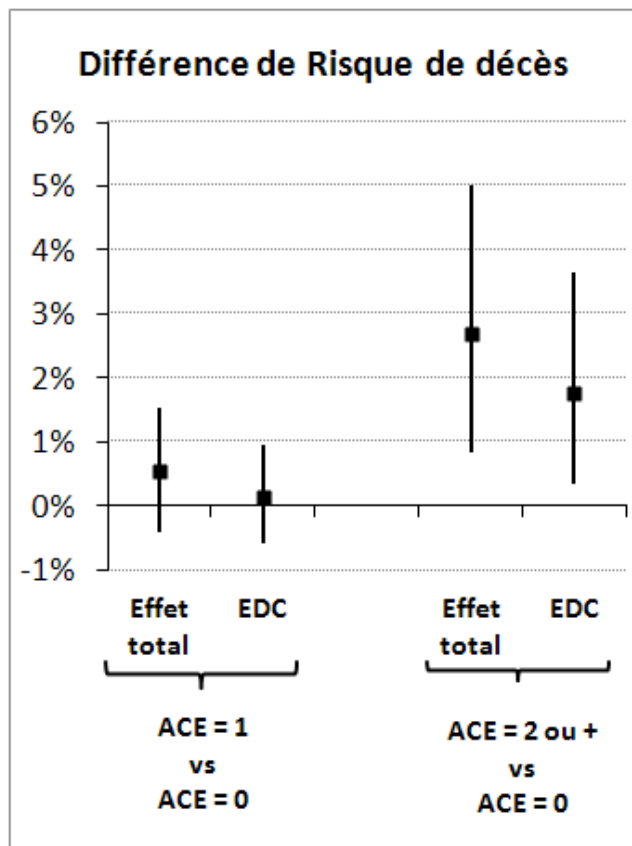
Estimation de l'effet direct contrôlé

Difficultés :

- **Choix de relations fonctionnelles entre les variables ?**
 - **Utiliser la totalité des vagues antérieures ?**
Vagues 3,4,5,6,7 pour les décès entre les vagues 7 et 8 !
 - **Modèles saturés ? (avec toutes les interactions)**
 - **Modèles « à périodes sensibles » ?**
différents coef de régression pour le tabac aux vagues 4, 5, 6...
 - **Modèles à risque cumulés ?**
Exemple du tabac en paquets-années

Résultats pour la cohorte NCDS 58

- Avec cette méthode, on estime que l'effet direct contrôlé n'est qu'en partie diminué par rapport à l'effet total de l'ACE sur la mortalité



Conclusions sur la g-computation

- **Intérêt de la g-computation**
 - Prise en compte des **facteurs de confusion intermédiaires, variables au cours du temps**
 - Possibilité d'intégrer les **interactions exposition * médiateurs**
 - Evaluer **plusieurs médiateurs**
- **Limites**
 - **Hypothèses de causalité** et hypothèses **fonctionnelles fortes**
 - Importance des **analyses de sensibilité** pour tester et comparer les résultats en fonction de différentes hypothèses
 - Analyses **longues** à réaliser

Merci de votre attention

Remerciements

- Financement : IRESP
- Center for Longitudinal Studies (CLS), Institute of Education, Economic and Social Data Service (ESDS) pour le recueil et la mise à disposition des données NCDS

Le CLS et l'ESDS n'ont aucune responsabilité dans les analyses et l'interprétation présentées ici

- Membres de la cohortes NCDS 58

Références sur la g-computation

- Taubman S et al. *Int J Epidemiol* 2009;38:1599-1611
- Daniel R et al. *Stat J* 2011;11:479-517.

Vague 3
16 ans

Vague 4
23 ans

Vague 5
33 ans

Vague 6
41/42 ans

Vague 7
46/47 ans

Vague 8
50 ans

Difficultés liées aux données manquantes +++

N initial = 12 164

% de données manquantes

19%

26%

26%

37%

0 à 16 ans

F 0

- sexe
- Scol parents
- cat soc parents
- poids naissance
- allaitement
- santé enfant

C3

Dépressivité
(16 ans)

C4

- Scolarité max
- santé perçue
- dépressivité
- handicap

C5

- cat sociale
- santé perçue
- dépressivité
- handicap

C6

- cat sociale
- santé perçue
- dépressivité
- handicap

C7

- cat sociale
- santé perçue
- handicap

ACE

M3

- tabac
- alcool
- obésité

M4

- tabac
- alcool
- obésité

M5

- tabac
- alcool
- obésité

M6

- tabac
- alcool
- obésité

M7

- tabac
- alcool

**Données sur les décès
≈ exhaustives
(pas de manquants)**

**Décès
3-4**

**Décès
4-5**

**Décès
5-6**

**Décès
6-7**

**Décès
7-8**

Vague 3
16 ans

Vague 4
23 ans

Vague 5
33 ans

Vague 6
41/42 ans

Vague 7
46/47 ans

Vague 8
50 ans

**Difficultés liées aux données
manquantes +++**

N initial = 12 164

Sous l'hypothèse MAR

- Appliquer une procédure d'imputation « simple » par équations en chaînes (van Buuren, 1999)
- Estimation des IC 95% par bootstrap \Rightarrow imputations « multiples »

On est plutôt MNAR

- Analyses de sensibilité :
ex : en majorant ou minorant les estimations moyennes par un coefficient choisi arbitrairement