

# Analyse longitudinale de trajectoires: approches multiniveaux paramétriques et non-paramétriques

Éric Lacourse

Département de Sociologie, Université de Montréal  
Centre de recherche de l'Hôpital Ste-Justine (GRIP)

Centre interuniversitaire québécois  
de statistiques sociales, Montréal, Québec

Le 21 octobre 2005

# Plan de la présentation

- Historique des études longitudinales
- Études longitudinales: l'état actuel des choses
- Avantages des modèles multiniveaux
- Le modèle multiniveau paramétrique
- Le modèle multiniveau non-paramétrique
- Extensions du modèle multiniveau non-paramétrique

# Historique

- Les premiers recensements périodiques, au niveau national, ont été réalisés en Nouvelle-France et par la suite au Québec de 1665 à 1754 (Menard, 2002)
- Aux États-unis, depuis la première décennie de l'existence officielle de la fédération américaine, on collecte des données longitudinales
- Les études longitudinales à long terme portant sur le développement des individus ont débutées après la première guerre mondiale
- Dans les années 70, les études longitudinales sont devenues plus populaires en sciences sociales
- Dans les années 90, on assiste à la plus grande croissance méthodologique avec le développement de logiciels d'analyse puissants (SAS, HLM, STATA)

# Définitions et terminologie

## ● Étude longitudinale

- Les données sont recueillies au moins à deux périodes pour chaque item, variable ou sujet
- Les sujets ou les cas analysés sont tous les mêmes ou du moins comparables d'une période à l'autre
- Les données sont rétrospectives ou prospectives

## ● Les multiples synonymes des analyses multiniveaux

- Paramétriques: random effects, growth curves, hierarchical linear model, ...
- Non-paramétriques: latent class, semiparametric mixture, latent class growth mixture model, ...

# Avantages et désavantages des études longitudinales

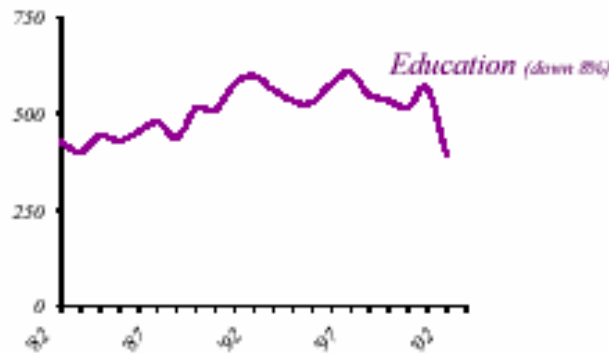
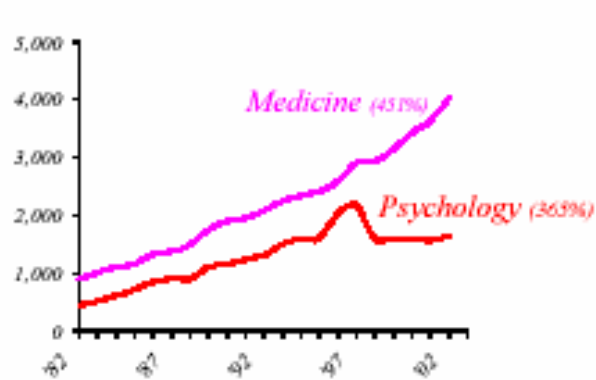
## ● Avantages

- Devis permettant de décrire des changements comportementaux et sociaux à travers le temps
- Devis permettant d'identifier les causes potentielles des phénomènes observés

## ● Désavantages

- La perte des sujets, changement dans l'équipe d'évaluation
- Coûts élevés en temps et en argent

Au cours des 20 dernières années, le nombre d'études longitudinales a augmenté très rapidement



# À quoi ressemblent ces études longitudinales?

La psychologie a été arbitrairement choisie comme thème et 10 articles relativement récents ont été sélectionnés aléatoirement ( parus en 1999 et 2003):

- ✓3 de *Developmental psychology*
- ✓3 de *Journal of Personality and Social Psychology*
- ✓2 de *Journal of Applied Psychology*
- ✓2 de *Journal of Consulting and Clinical Psychology*



Nous dénombrons plus de 150 articles par année dont plusieurs comportent des études longitudinales:

- ✓ En 1999, 33 % d'études longitudinales
- ✓ En 2003, 47 % d'études longitudinales

## La bonne nouvelle:

Un pourcentage croissant des études dites longitudinales correspond à des études qui sont réellement longitudinales (i.e., elles comportent plus de 2 vagues de mesure)



	1999	2003
2 vagues de mesure	36%	26%
3 vagues de mesure	26%	29%
4 vagues de mesure	38%	45%

# Une partie du problème pourrait être l'ignorance des évaluateurs

*Ex: Commentaires reçus en 2003 critiquant un article portant sur le modèle de croissance individuelle défini sur 3 vagues de données.*

## Évaluateur A

« Je ne comprend pas les statistiques de cette étude suffisamment pour évaluer leur justesse. J'imagine que 99% des lecteurs de *Developmental Psychology* se trouvent dans le même cas. ... Des études précédentes dans ce domaine ont utilisé des méthodes simples de corrélation ou de régression qui produisent des résultats plus faciles à interpréter. ... Les auteurs sont félicités pour leur étude longitudinale détaillée mais les statistiques ne sont pas assez intuitives. ... Je trouve donc que cet article n'a pas vraiment sa place dans *Developmental Psychology*. »

## Évaluateur B

« L'analyse ne réussit pas à mettre en lumière les éléments amorcés dans une introduction claire et convaincante. ... Je dois reconnaître que j'ai débuté dans ce champ avant l'arrivée des méthodes sophistiquées ainsi qu'avant l'ère des modèles de croissance et que ceux-ci ont toujours, dans une certaine mesure, piqué ma curiosité. J'ai tenté de rester ouvert, il n'en demeure pas moins que certaines de mes observations peuvent s'avérer naïves, voir même inappropriées. »

# Avantages spécifiques aux approches multiniveaux

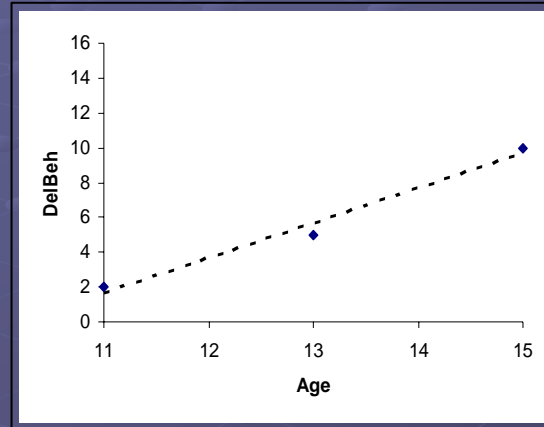
- Elles accommodent bien les données continues, catégorielles et même celles qui sont manquantes
- Elles permettent l'introduction de pondérations pour les échantillons complexes
- Elles ne nécessitent pas que les intervalles de temps soient équivalents
- Elles accommodent les échantillons à cohortes multiples
- Elles permettent de dériver des intervalles de confiance pour l'ensemble des paramètres (méthode de bootstrap)

# Le modèle multiniveau paramétrique

# Modèle du processus de changement: Vue d'ensemble

Assumons l'existence d'un modèle statistique hiérarchique à deux niveaux

**Au niveau 1:** Le modèle présente la trajectoire individuelle de changement qui démontre comment l'état de chaque sujet dépend du temps.



Ex: différence dans les comportements délinquants des adolescents selon le sexe.

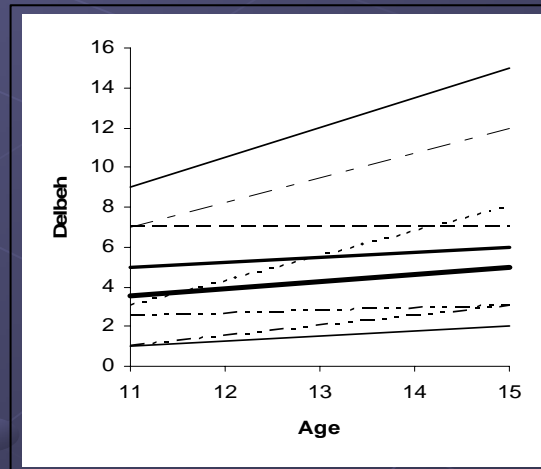
$$Y_{ij} = [\pi_{0i} + \pi_{1i}(\text{AGE}_{ij}) + \varepsilon_{ij}]$$

L'ordonnée à l'origine pour un sujet i

La pente pour un sujet i

Résidu pour un sujet i à chaque période j

**Au niveau 2:** Le modèle présente les différences interindividuelles du changement. Décrivant comment les trajectoires des caractéristiques Individuelles (i.e. l'ordonnée à l'origine et la pente) varient d'un sujet à l'autre.



Modèle de niveau 2 pour les ordonnées à l'origine du modèle de niveau 1 :

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\text{MALE} + \zeta_{0i}$$

Modèle de niveau 2 pour les pentes du modèle de niveau 1 :

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\text{MALE} + \zeta_{1i}$$

# Le changement dans la consommation d'alcool des adolescents en fonction de l'alcoolisme des parents

• *Question de recherche* : Est-ce que les trajectoires individuelles de consommation d'alcool durant l'adolescence sont influencées par: 1- l'alcoolisme des parents, 2- et la consommation d'alcool des pairs?

• *Source*: Curran, Stice, et Chassin (1997)

• *Échantillon*: 82 adolescents

• *Déroulement*:

– 3 vagues balancées à intervalles égaux : à l'âge de 14, 15 et 16 ans les adolescents indiquent leur consommation d'alcool au cours de la dernière année

– La consommation d'alcool a été évaluée par 4 items en utilisant une échelle à 8 niveaux (boire de la bière ou du vin, boire de l'alcool fort, boire 5 consommations ou plus d'un coup, et être soûl). Chaque item obtient un score sur une échelle de 8 points (0="pas du tout" à 7="tous les jours") pour un maximum de 32 points

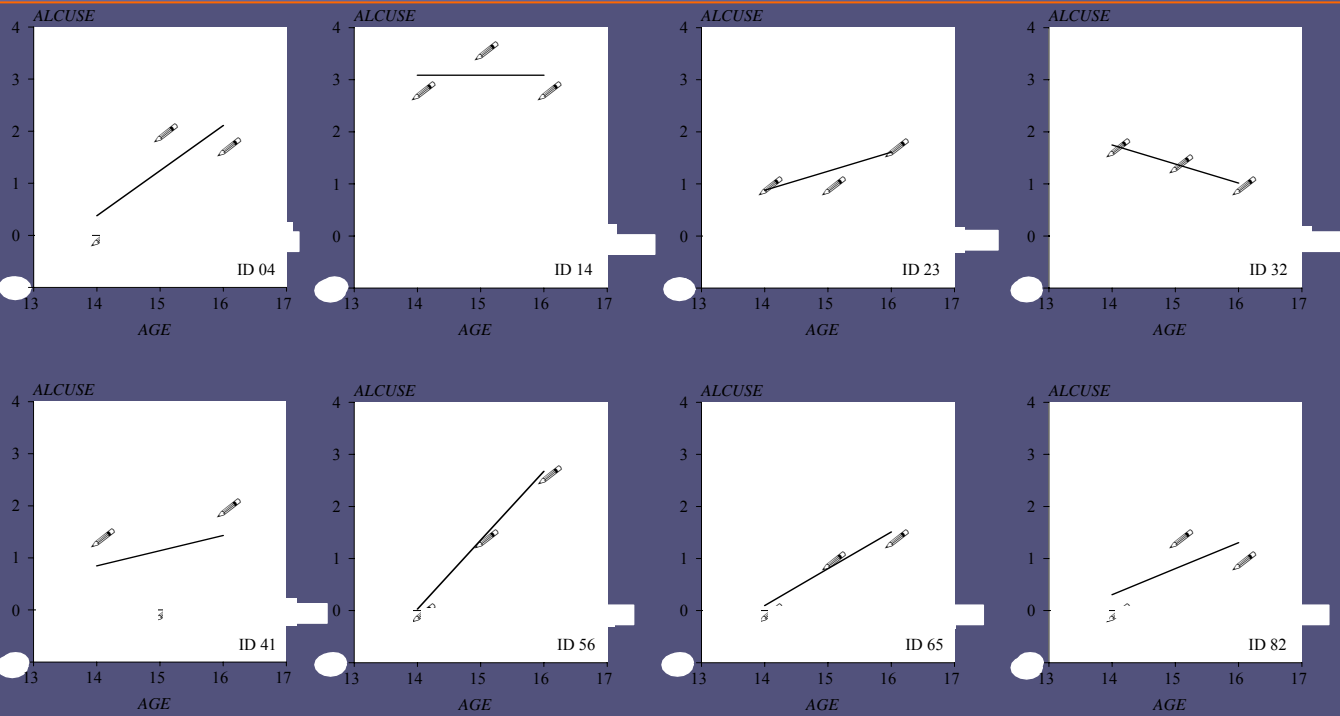
• *Les variables analysées*

– **ALCUSE**: Est la racine carré de la somme des 4 items

– **COA**: L'adolescent est-il l'enfant d'un parent alcoolique?

– **PEER**: Une mesure de la consommation d'alcool par les pairs à l'âge de 14 ans

# Quelle serait la forme idéale de la trajectoire de niveau-1?



## 3 caractéristiques de ces tracés

1. *Approximativement linéaires, souvent (mais pas tout le temps) en croissance.*
2. *certaines trajectoires OLS s'ajustent bien.*
3. *Certaines trajectoires OLS affichent plus de discontinuité.*

$$ALCUSE_{ij} = \pi_{0i} + \pi_{1i} (AGE_{ij} - 14) + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$$

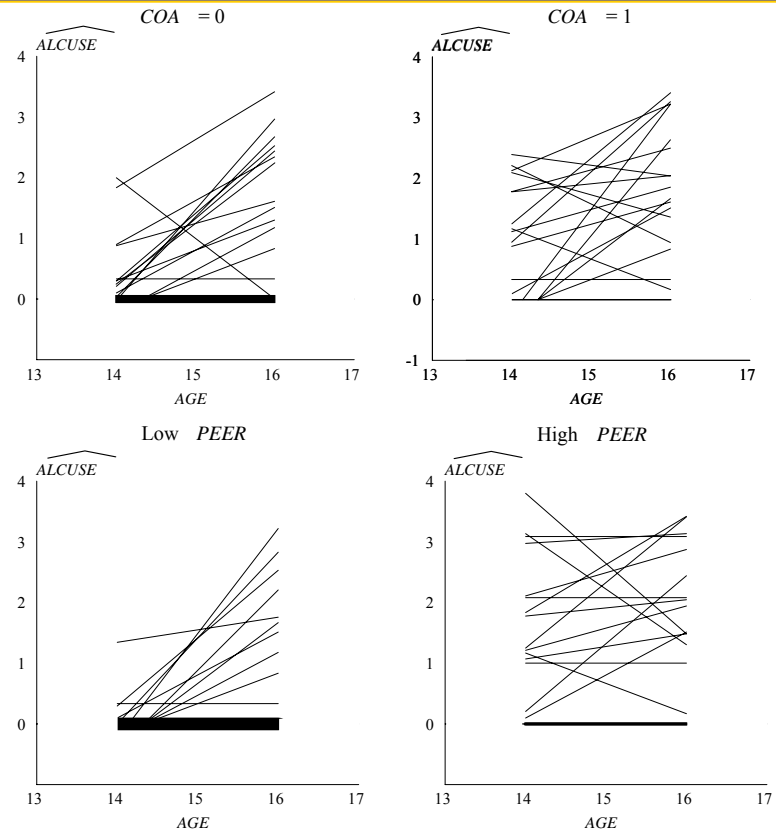
$$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij}$$

Statut initial réel  
(ie, quand TIME=0)

Taux de changement  
réel par unité de temps  
(TIME)

Proportion des résultats  
inexpliqués pour le sujet i à  
l'occasion j

# À quoi ressemblent les sous modèles de niveau-2 pour l'évaluation des différences du processus de changement?



**Niveau 2: ordonnée à l'origine**  
 La moyenne du statut initial et le taux de changement de la population pour les non-COA

**Niveau 2: pentes**  
 L'effet d'être COA sur le statut initial et le taux de changement

$$\pi_{0i} = \gamma_{00} + \gamma_{01} COA_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} COA_i + \zeta_{1i}$$

Examinons les variations dans les trajectoires OLS de niveau-1 par:

- COA:** Les COA ont des ordonnées à l'origine plus élevées mais des pentes moins abruptes
- PEER (cut-off à la moyenne):** les adolescents qui ont des amis qui boivent à 14 ans ont des ordonnées à l'origine plus élevées mais des pentes plus près de zéro et mêmes déclinantes

**Résidus du modèle de niveau-2**  
 Déviations des trajectoires individuelles autour des moyennes prédites

$$\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

# Comment évaluer l'effet de la variable COA avec une variable contrôle?

		Parameter	Model C	Model D
<b>Fixed Effects</b>				
Initial status, $\pi_{0i}$	Intercept	$\gamma_{00}$	0.316*** (0.131)	-0.317*** (0.148)
	COA	$\gamma_{01}$	0.743*** (0.195)	0.579*** (0.162)
	PEER	$\gamma_{02}$		0.694*** (0.112)
Rate of change, $\pi_{2i}$	Intercept	$\gamma_{10}$	0.293*** (0.084)	0.429*** (0.114)
	COA	$\gamma_{11}$	-0.049 (0.125)	-0.014 (0.125)
	PEER	$\gamma_{12}$		-0.150 (0.086)
<b>Variance Components</b>				
Level 1	Within-person	$\sigma_{\epsilon}^2$	0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	$\sigma_0^2$	0.488** (0.128)	0.241** (0.093)
	In rate of change	$\sigma_1^2$	0.151* (0.056)	0.139* (0.055)
	Covariance	$\sigma_{01}$	-0.059 (0.066)	-0.006 (0.055)
<b>Pseudo R<sup>2</sup> Statistics and Goodness-of-fit</b>				
			.150	.291
	$R_{yy}^2$		.40	.40
	$R_{\epsilon}^2$		.218	.614
	$R_0^2$		.000	.079
	$R_1^2$		621.2	588.7
	Deviance		637.2	608.7
	AIC		656.5	632.8
	BIC			

~p < .10; \*p < .05; \*\*p < .01; \*\*\*p < .001

## L'effet de COA

- L'estimé différentiel pour ALCUSE entre les COA et les non-COA, en contrôlant pour PEER, est de 0.579 ( $p < .001$ )
- Pas de différence significative pour le taux de changement.

## L'effet de PEER

- Les adolescents qui ont des pairs qui boivent plus à 14 ans, boivent eux aussi plus à 14 ans (statut initial).
- Effet négatif modeste sur le taux de changement ( $p < .10$ ).

- Pas différent (comme escompté)

- Pris ensemble, PEER et COA expliquent:
  - 61.4% de la variance du statut initial,
  - 7.9% de la variance pour le taux de changement.

Étape suivante?

Si nous avons d'autres variables prédictives nous les ajouterions parce que les composantes de la variance sont encore statistiquement significatives

**Simplifions le modèle?**  
Puisque la variable COA n'est pas associée au taux de changement, Pouvons nous l'enlever du modèle?

# Comment représenter les résultats? Les tracés prototypiques

Commençons simplement par interpréter le *Modèle C*, qui contient les effets principaux de COA autant sur le *statut initial* que sur le *taux de changement* au niveau 2

$$\bar{\pi}_{0i} = 0.316 + 0.743 \text{ COA}_i$$

$$\bar{\pi}_{1i} = 0.293 - 0.049 \text{ COA}_i$$

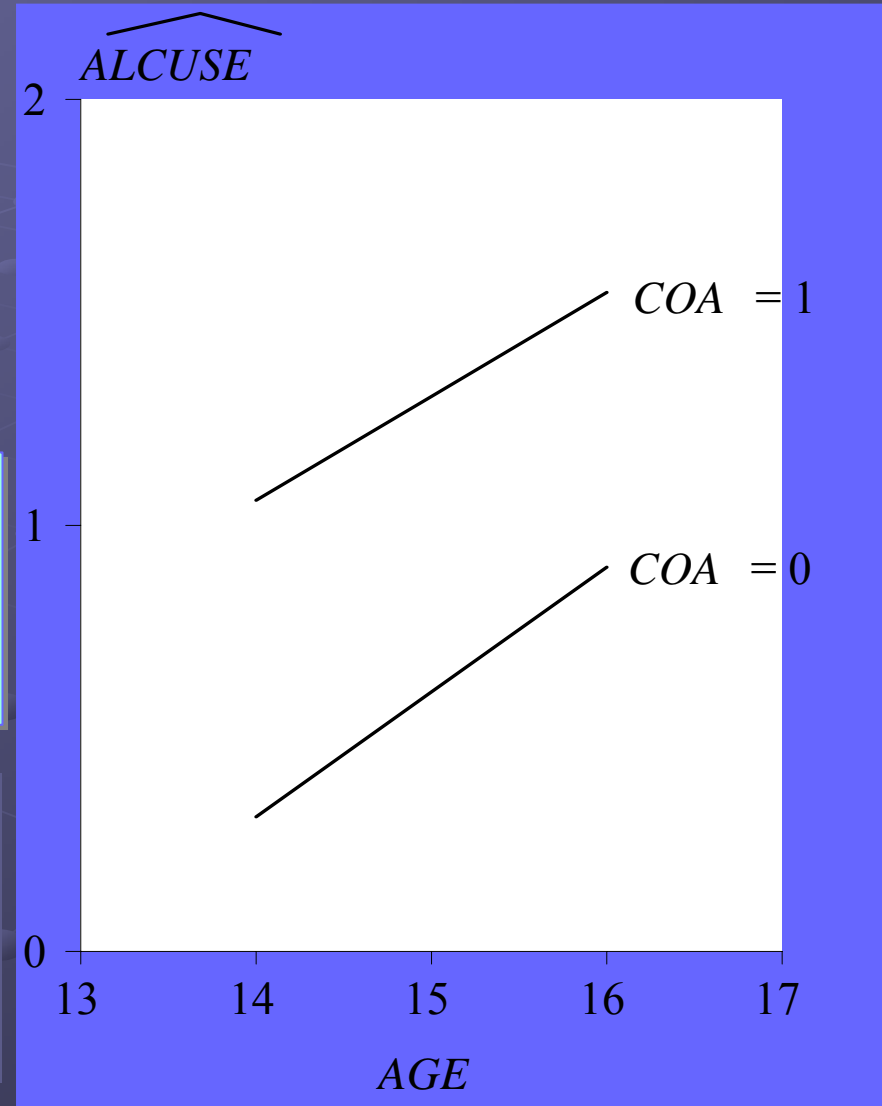
Substituons les valeurs observées de COA (0 et 1)

$$\begin{array}{l} \text{When } \text{COA}_i = 0 \\ \text{When } \text{COA}_i = 1 \end{array} \left\{ \begin{array}{l} \bar{\pi}_{0i} = 0.316 + 0.743(0) = 0.316 \\ \bar{\pi}_{1i} = 0.293 - 0.049(0) = 0.293 \\ \bar{\pi}_{0i} = 0.316 + 0.743(1) = 1.059 \\ \bar{\pi}_{1i} = 0.293 - 0.049(1) = 0.244 \end{array} \right.$$

Substituons les paramètres de croissance estimés au modèle de croissance individuelle:

$$\text{When } \text{COA}_i = 1: \hat{Y}_{ij} = 1.059 + 0.244 \text{ TIME}_{ij}$$

$$\text{When } \text{COA}_i = 0: \hat{Y}_{ij} = 0.316 + 0.293 \text{ TIME}_{ij}$$



Tracé des trajectoires de croissance prototypiques

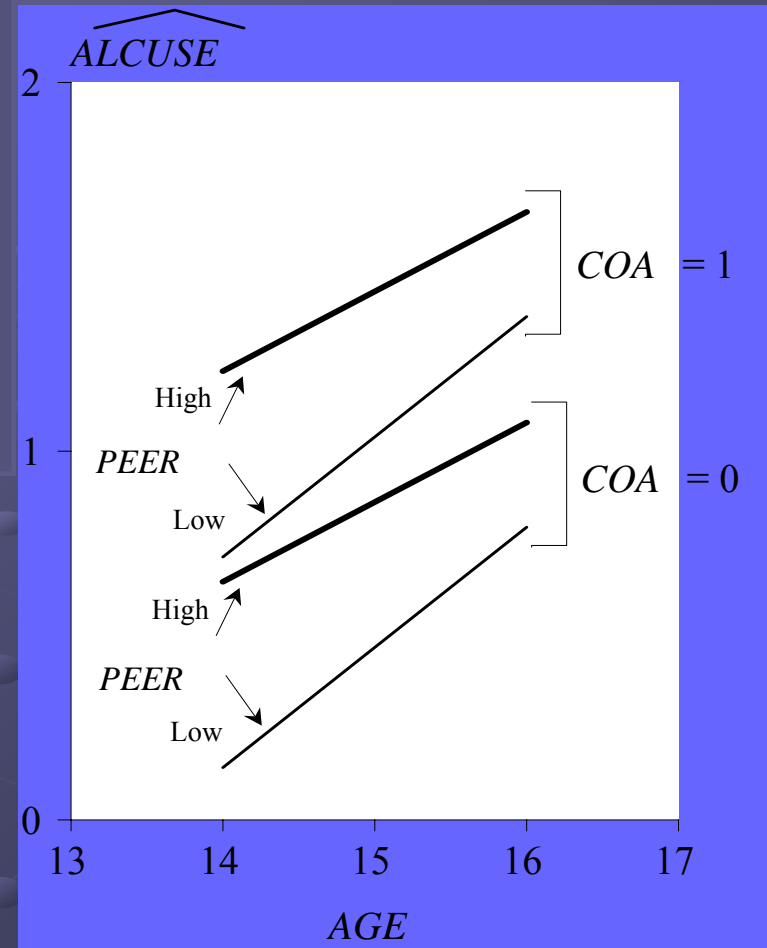
# Comment représenter les trajectoires prototypiques de changement quand les variables prédictives ne sont pas toutes dichotomiques?

Substituons les valeurs intéressantes des estimateurs au modèle et traçons les trajectoires prototypes:

- Choisissez des valeurs pertinentes (ex., 12, & 16 ans de scolarité).
- Utilisez une gamme sensée de percentiles (e.g., 10<sup>th</sup>/50<sup>th</sup>/90<sup>th</sup>).
- Utilisez la moyenne simple  $\pm .5$  (ou 1) écart-type.
- Utilisez seulement la moyenne simple si vous voulez simplement contrôler pour l'impact d'une variable prédictive au lieu de représenter son effet.

Interprétons le Modèle E, qui contient l'effet principal de COA et PEER., pour l'échantillon.

$$\begin{aligned} \overline{PEER} &= 1.018, \quad sd_{PEER} = 0.726 \\ \text{Low PEER} &= 1.018 - .5(0.726) = 0.655 \\ \text{High PEER} &= 1.018 + .5(0.726) = 1.381 \end{aligned}$$



PEER	COA	Initial status ( $\hat{\pi}_{0i}$ )	Rate of change ( $\hat{\pi}_{1i}$ )
Low	No	$-0.314 + 0.695(0.655) + 0.571(0) = 0.142$	$0.425 - 0.151(0.655) = 0.326$
Low	Yes	$-0.314 + 0.695(0.655) + 0.571(1) = 0.713$	$0.425 - 0.151(0.655) = 0.326$
High	No	$-0.314 + 0.695(1.381) + 0.571(0) = 0.646$	$0.425 - 0.151(1.381) = 0.216$
High	Yes	$-0.314 + 0.695(1.381) + 0.571(1) = 1.217$	$0.425 - 0.151(1.381) = 0.216$

# Comment utiliser la déviance pour évaluer simultanément les effets de l'ajout de variables prédictives aux deux modèles de niveau 2?

Comparons le **Modèle C** (qui inclut les effets de COA autant pour le statut initial que pour le taux de changement) au **Modèle B**.

			Parameter	Model B	Model C
<b>Fixed Effects</b>					
Initial status, $\pi_{0i}$	Intercept	$\gamma_{00}$		0.651*** (0.105)	0.316*** (0.131)
	COA	$\gamma_{01}$			0.743*** (0.195)
	PEER	$\gamma_{02}$			
Rate of change, $\pi_{2i}$	Intercept	$\gamma_{10}$		0.271*** (0.062)	0.293*** (0.084)
	COA	$\gamma_{11}$			-0.049 (0.125)
	PEER	$\gamma_{12}$			
<b>Variance Components</b>					
Level 1	Within-person	$\sigma_{\varepsilon}^2$		0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	$\sigma_0^2$		0.624*** (0.148)	0.488** (0.128)
	In rate of change	$\sigma_1^2$		0.151** (0.056)	0.151* (0.056)
	Covariance	$\sigma_{01}$		-0.068 (0.070)	-0.059 (0.066)
<b>Pseudo R<sup>2</sup> Statistics and Goodness-of-fit</b>					
	$R_{xy}^2$			.043	.150
	$R_{\varepsilon}^2$			.40	.40
	$R_0^2$				.218
	$R_1^2$				.000
	Deviance			636.6	621.2
	AIC			648.6	637.2
	BIC			663.0	656.5

Nous obtenons le **Modèle B** à partir du **Modèle C** en ajoutant 2 contraintes indépendantes (seulement sur l'effet fixe):

$$H_0: \gamma_{01}=0, \gamma_{11}=0$$

## Différence dans la déviance

$$636.61 - 621.20 = 15.41 \quad (2 \text{ df}, p < .001)$$

Le modèle de croissance conditionnelle procure un meilleur ajustement que le modèle de croissance inconditionnelle.

Notons que le test n'implique pas que chaque pente de niveau 2 soit elle-même statistiquement significative.

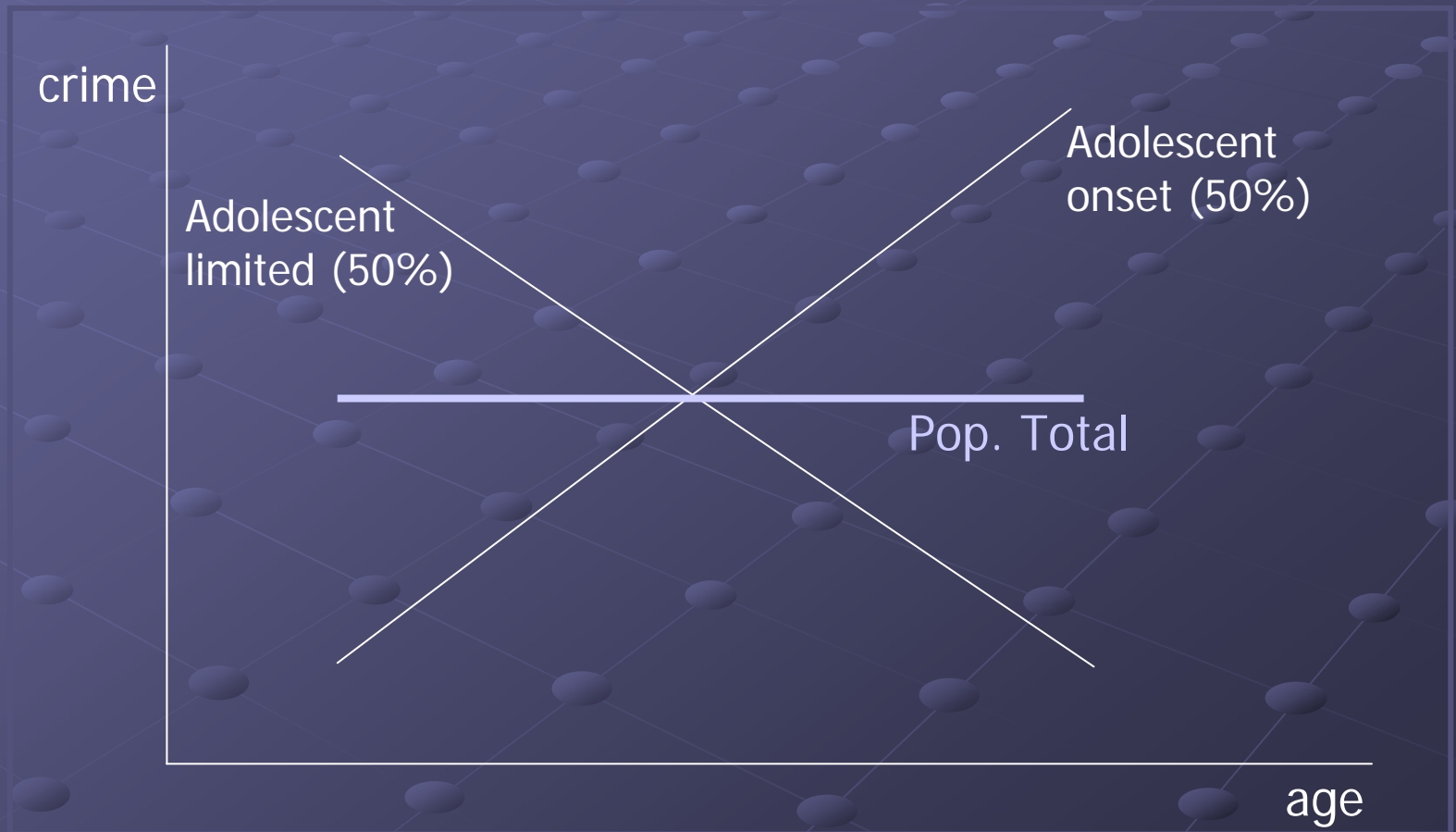
# Le modèle multiniveau non-paramétrique

# Théorie portant sur les trajectoires développementales discrètes



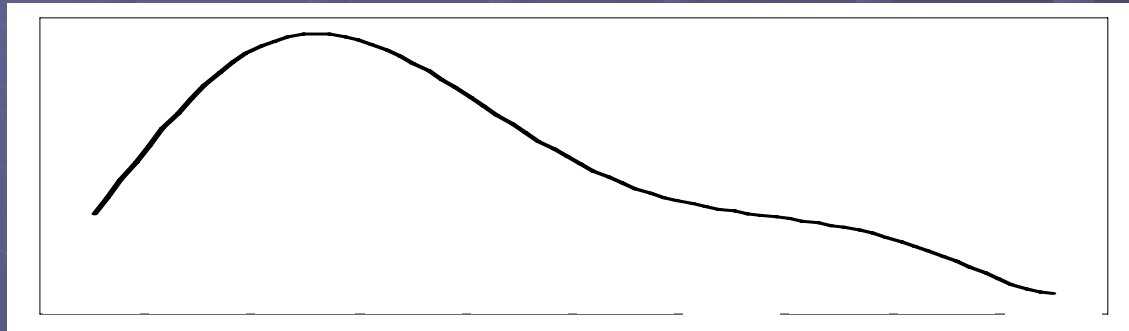
- Hétérogénéité qualitative dans le développement des trajectoires de délinquance
  - Moffitt (1993) & Patterson (1989) Early onset vs. Late onset

# Identification de trajectoires développementales distinctes: Une illustration

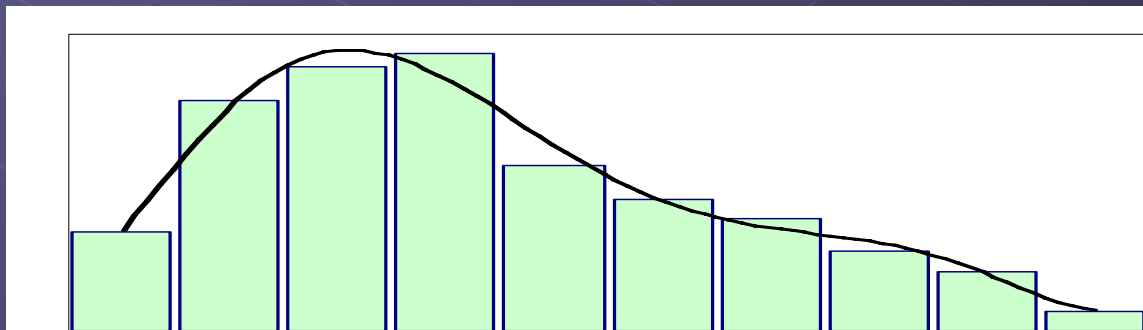


# L'utilisation de groupes pour permettre l'approximation d'une distribution inconnue

Panel A



Panel B



# Pourquoi une approximation?

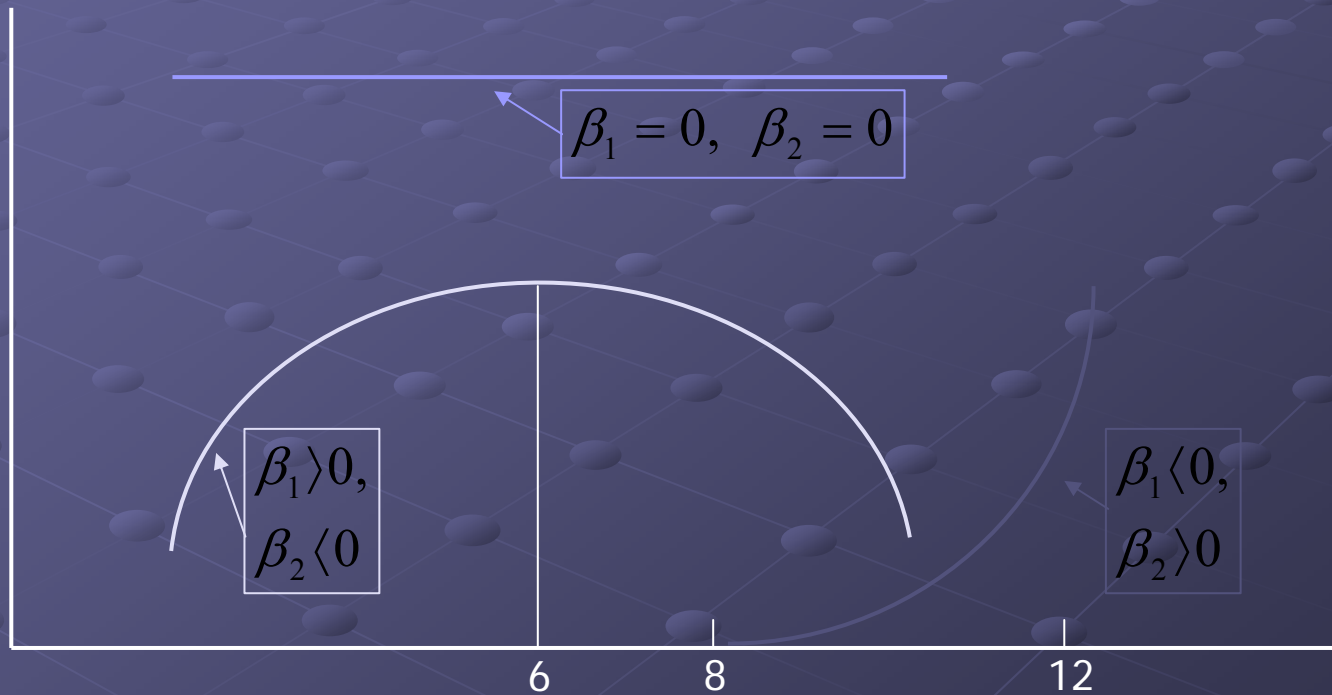
## Parce que tous les modèles sont des approximations

“un modèle bien fait n'est, après tout, qu'une série de mensonges judicieusement choisis, ou plus précisément, de vérités partielles concernant une réalité. Ce modèle a été choisi parce qu'il nous permet de concevoir un problème d'une manière plus appropriée qu'en utilisant un autre modèle

William Baumol (1997:55)

# Le lien entre l'âge et le comportement

$$\log(\lambda_{it}^j) = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2 + \beta_3^j \text{Age}_{it}^3$$

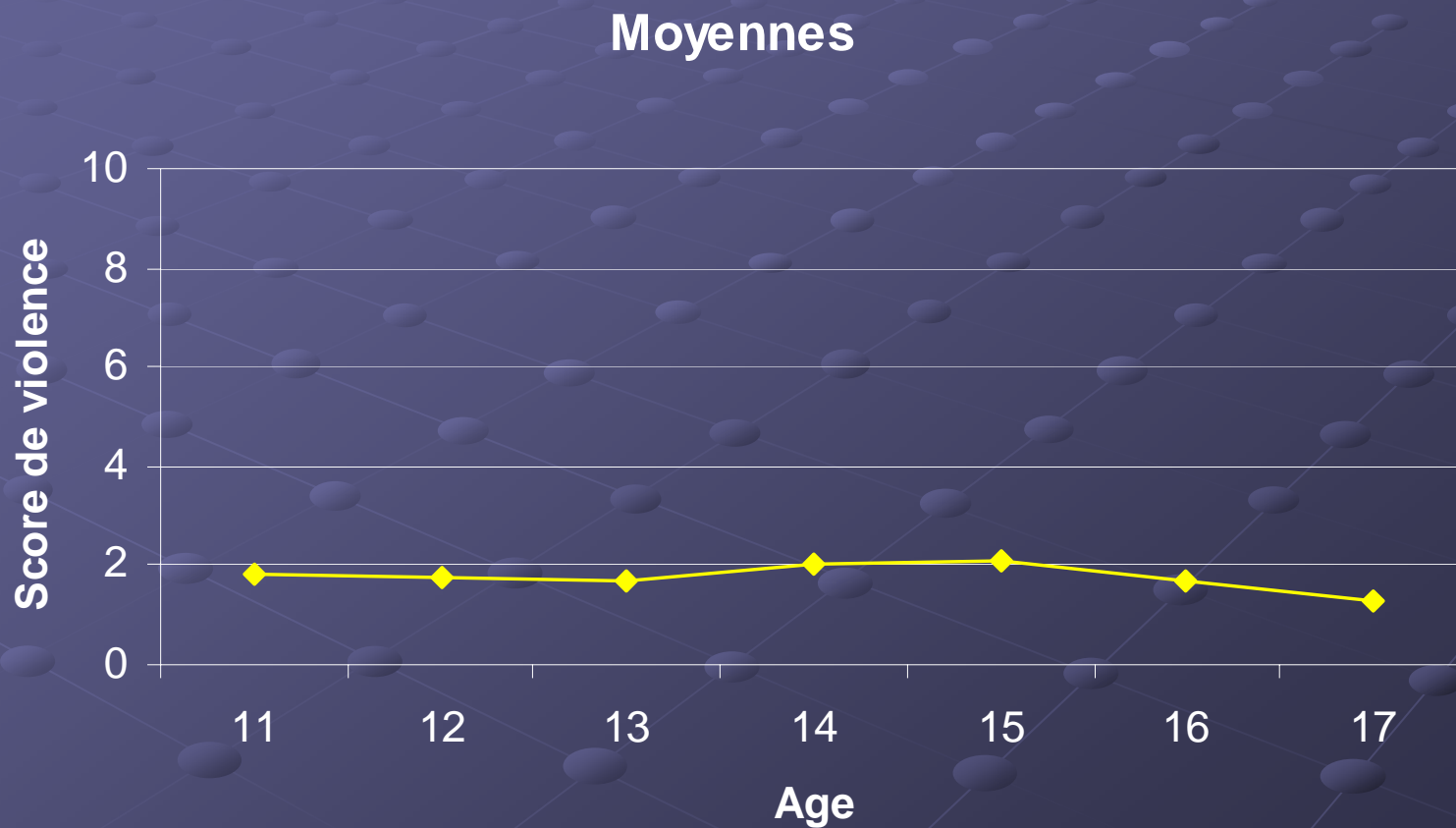


$$\beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2$$

# Avantages de l'approche méthodologique par regroupement de trajectoires

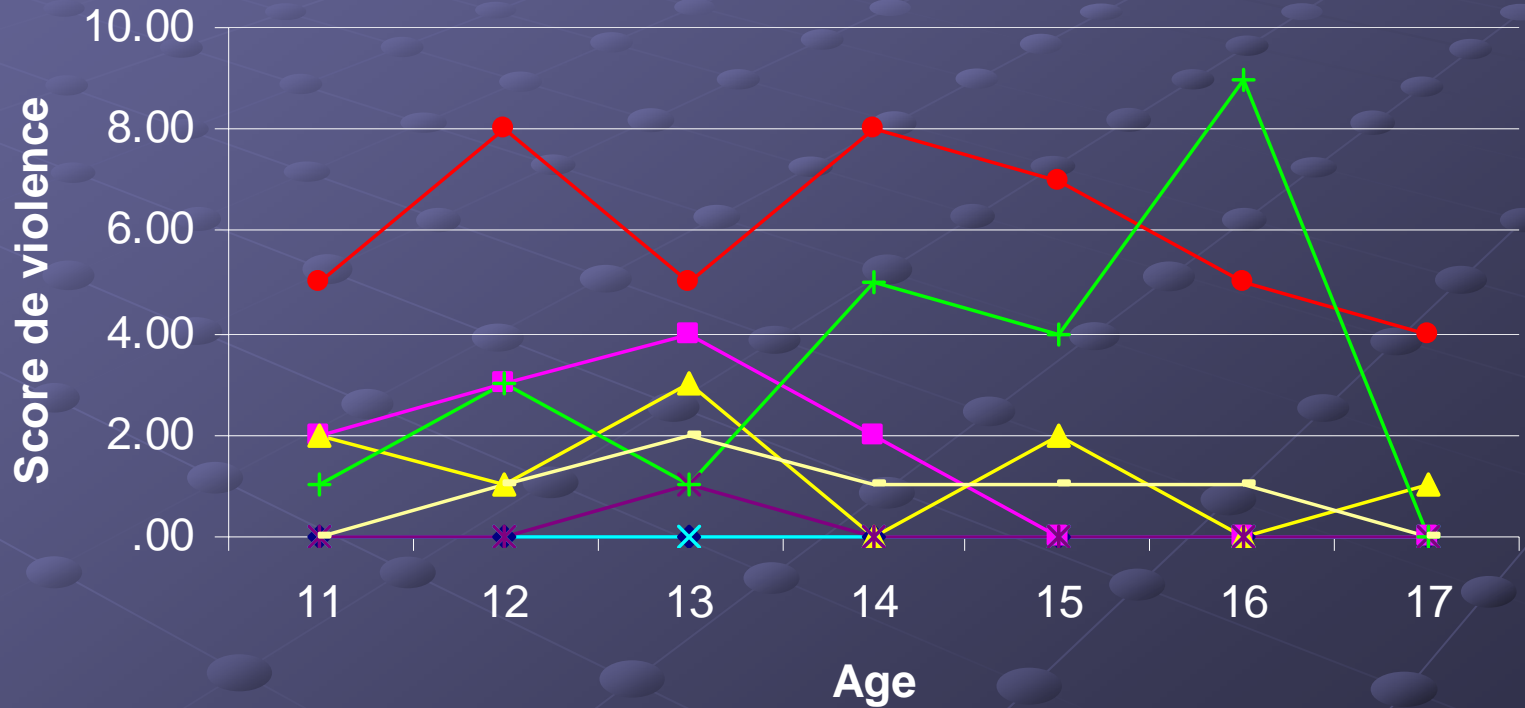
- Elle identifie au lieu de présumer la présence de trajectoires développementales distinctes
- Elle estime la proportion de la population suivant une trajectoire spécifique
- Elle permet d'identifier les caractéristiques particulières à chaque regroupement de trajectoires

# Moyennes des scores de violence durant l'adolescence

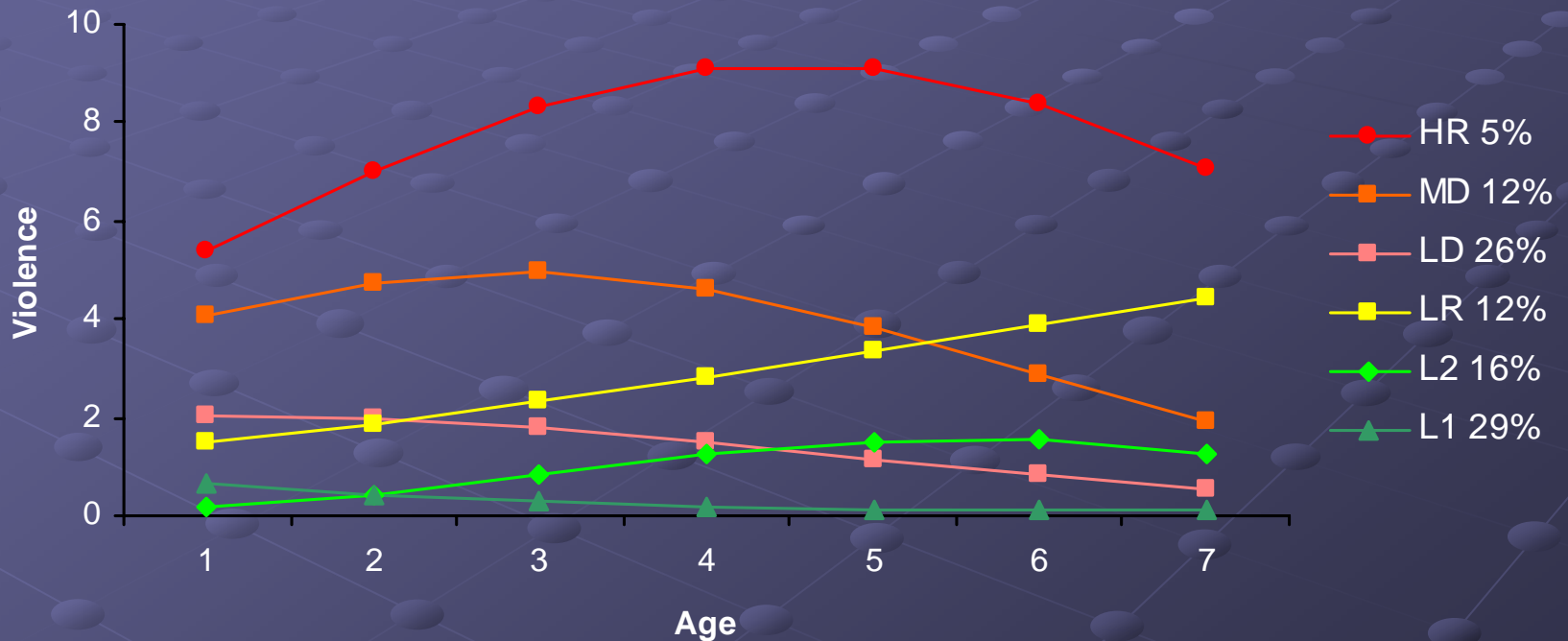


# Trajectoires individuelles de 10 enfants participant à l'étude

## Trajectoires individuelles



# Trajectoires développementales des comportements violents

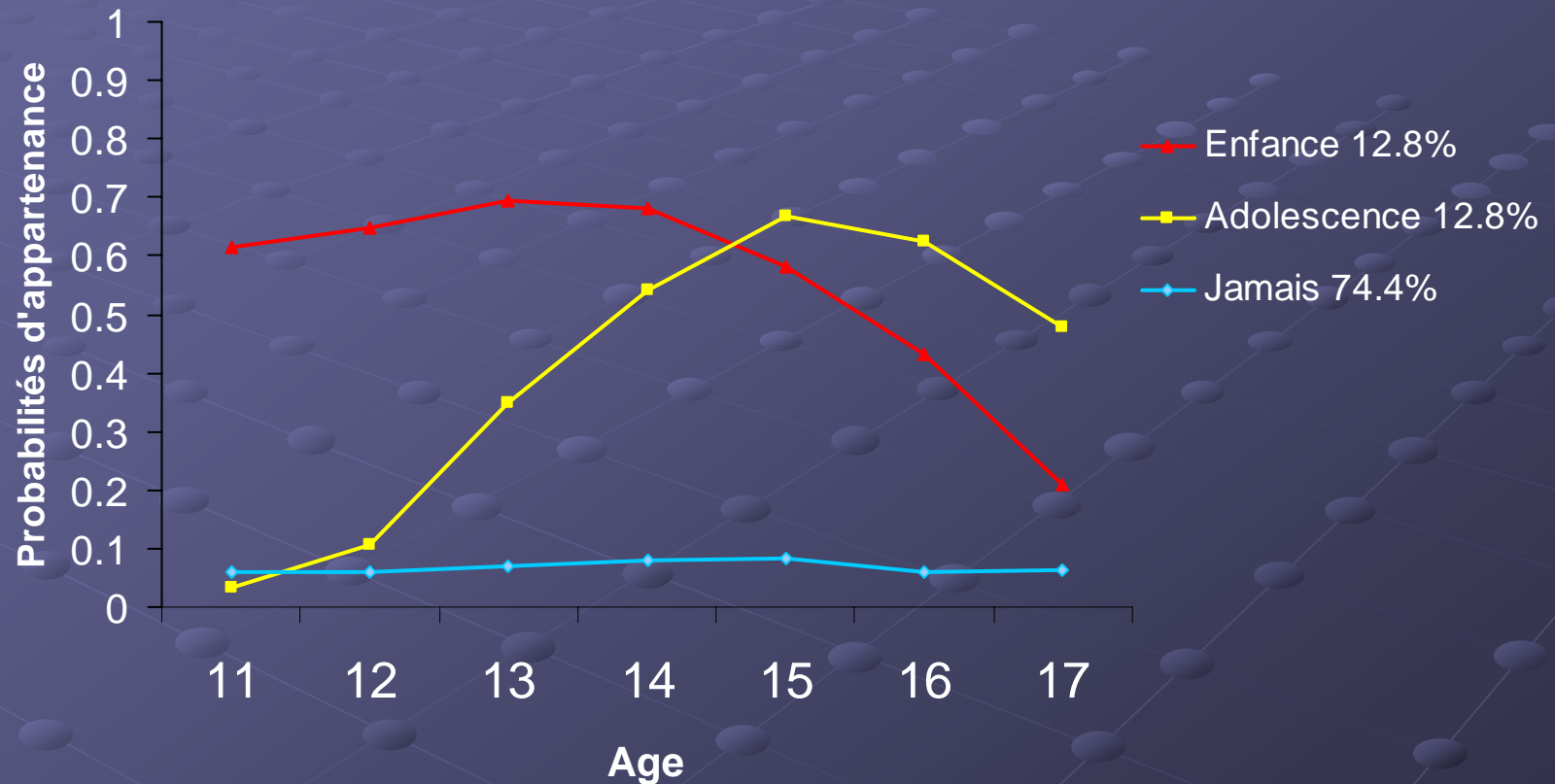


Lacourse, E., Côté, S., Nagin, D. S., Tremblay, R. E., Vitaro, F., & Brendgen, M. (2002). A longitudinal-experimental design to test developmental theories of antisocial behavior. *Development and Psychopathology*, 14, 909-924

# Étapes recommandées dans la sélection du modèle final

- Il faut décider du nombre de trajectoires et de leurs formes
  - Modéliser des courbes quadratiques au départ et ajouter un groupe à la fois
  - Utiliser le BIC (si possible)
  - Raffiner la forme des trajectoires en fonction de la signification statistique des paramètres estimés
  - Utiliser les diagnostics d'ajustement (probabilités postérieures)

# Trajectoires d'appartenance à un groupe de pairs déviants



Lacourse, E., Nagin, D., Tremblay, R. E., Vitaro, F., & Claes, M. (2003). Developmental trajectories of boys delinquent group membership and facilitation of violent behaviors during adolescence. *Development and Psychopathology*, 15, 183-197.

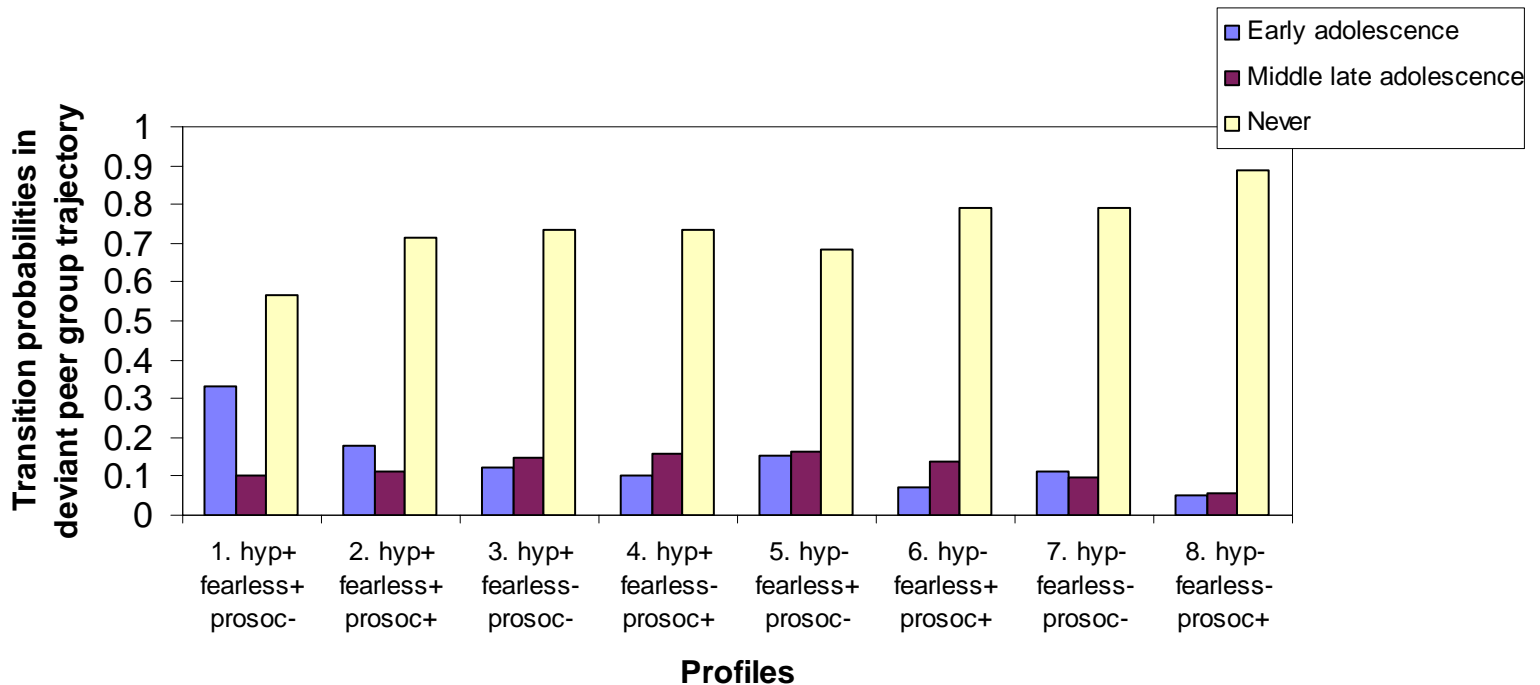
# Établir les liens entre les caractéristiques individuelles et les trajectoires

- Faire des comparaisons univariées au départ
- Introduire les covariables simultanément pour considérer l'incertitude dans la classification
  - Les liens multivariés sont faits à l'aide d'une régression logistique multinomiale

# Exemple de facteurs dans l'enfance associés à l'affiliation à des pairs déviants

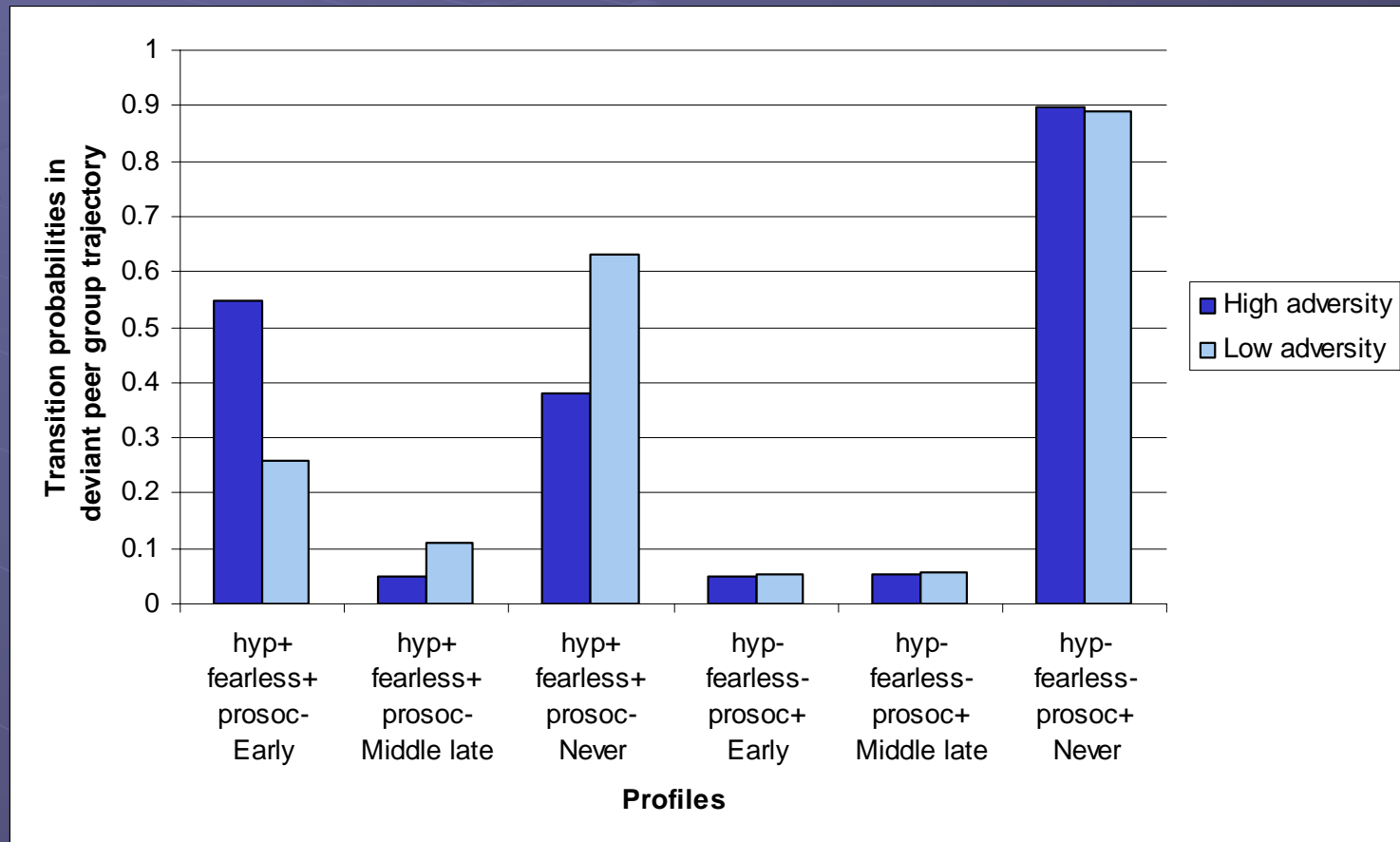
- Hyperactivité
- Prosocialité
- Anxiété
- Adversité familiale

# Probabilités de transition au sein des trois trajectoires conditionnelles aux profils comportementaux en maternelle



Lacourse, E., Nagin, D., Vitaro, F., Côté, S., Arseneault, L. & Tremblay, R. E., (in press). Prediction of early onset deviant peer group affiliation: a 12 year longitudinal study. Archives of General Psychiatry

# Probabilités de transition au sein des trois trajectoires conditionnelles aux profils comportementaux en maternelle (suite)

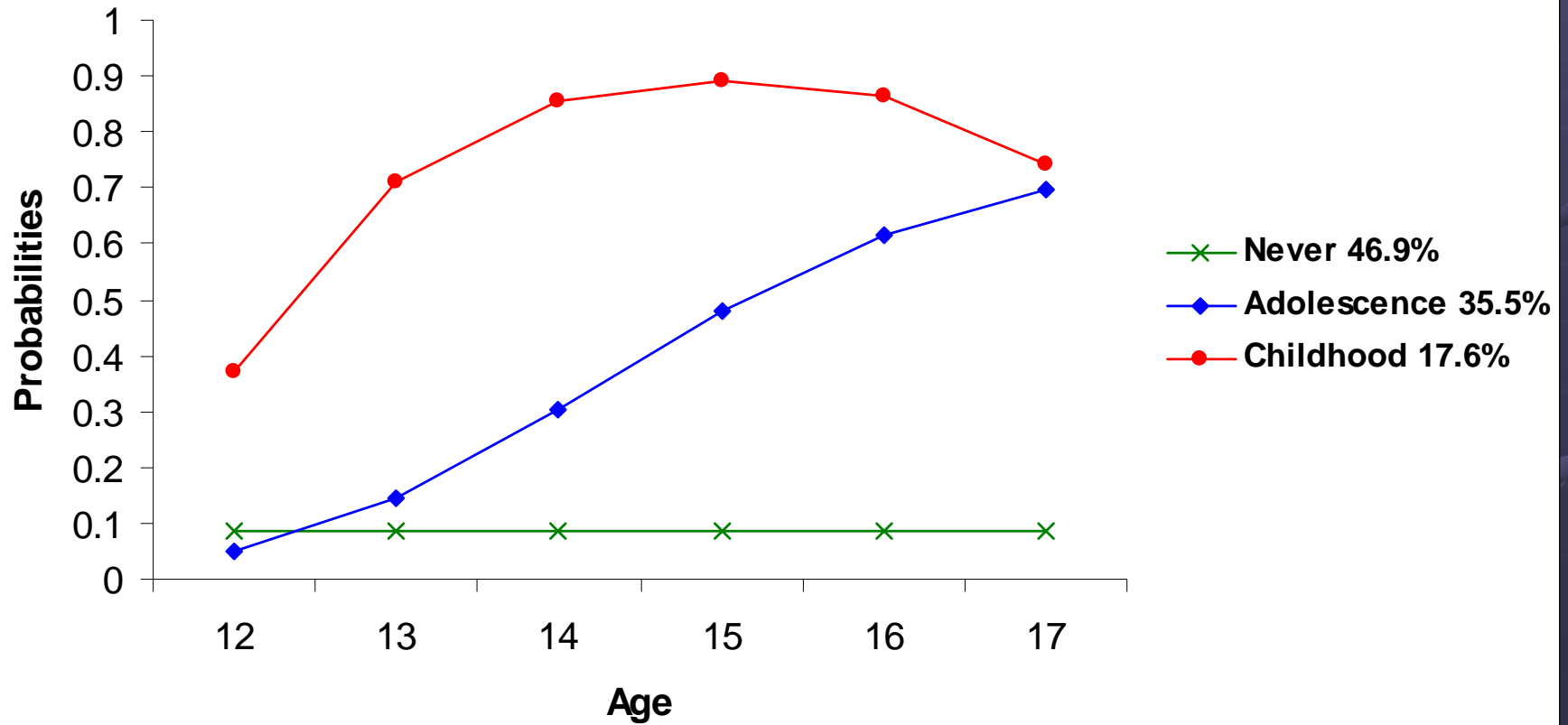


Lacourse, E., Nagin, D., Vitaro, F., Côté, S., Arseneault, L. & Tremblay, R. E., (in press). Prediction of early onset deviant peer group affiliation: a 12 year longitudinal study. Archives of General Psychiatry

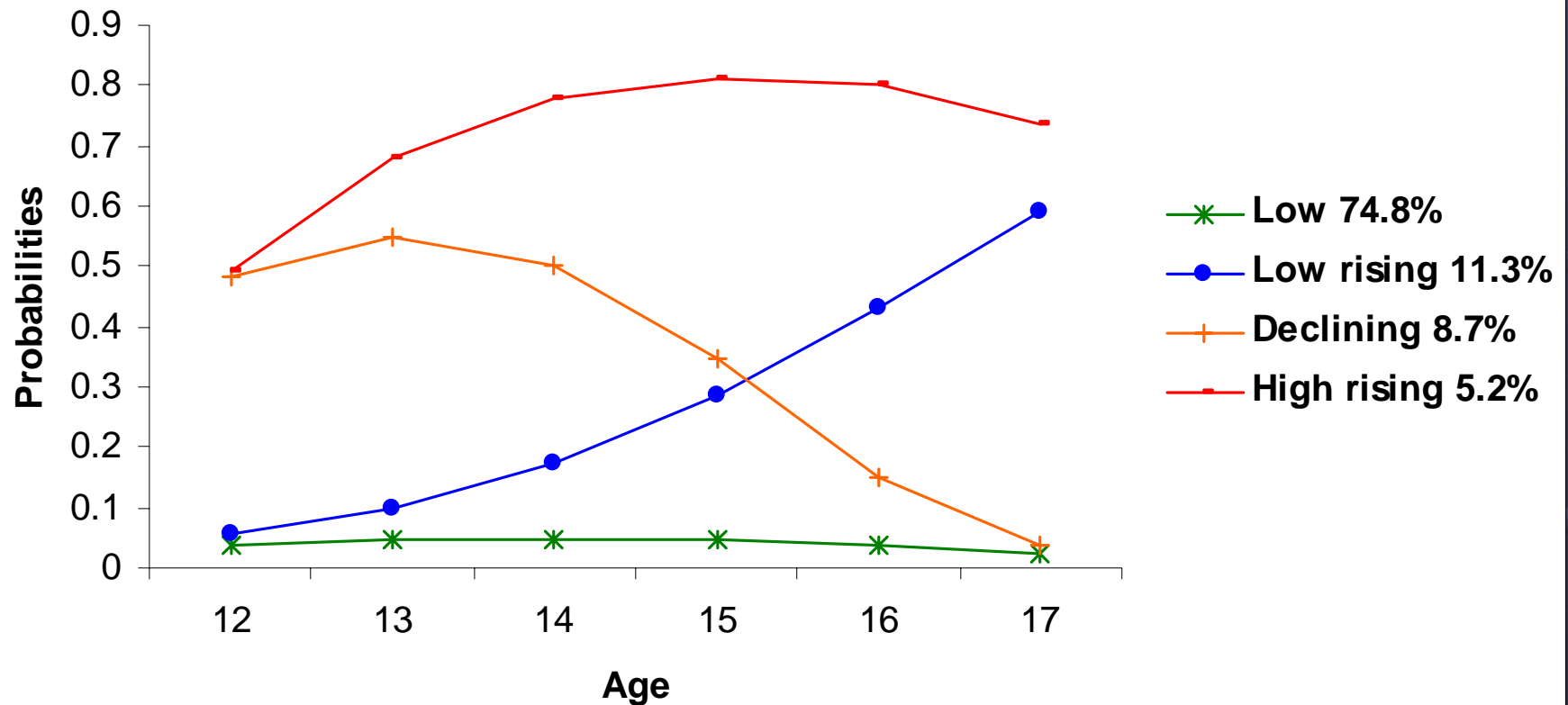
# Extensions du modèle multiniveau non-paramétrique

- Analyser des trajectoires jointes
- Insérer des covariables à l'intérieur des trajectoires

# Trajectoires d'appartenance à un groupe de pairs déviants (amis arrêtés par la police)



# Trajectoires d'agression avec une arme



# Trajectoires d'agression avec une arme conditionnelles aux trajectoires de pairs déviants

Deviant peers Group	Weapon use Group			
	Low	Low rising	Declining	High rising
Never	<b>.98</b>	.08	.09	.00
Adolescence	.51	<b>.39</b>	.10	.00
Childhood	.18	.13	<b>.40</b>	<b>.29</b>

# Trajectoires de pairs déviants conditionnelles aux trajectoires d'agression avec une arme

Deviant peers Group	Weapon use Group			
	Low	Low rising	Declining	High rising
Never	<b>.65</b>	.02	.03	.00
Adolescence	.29	<b>.83</b>	.31	.00
Childhood	.06	.15	<b>.66</b>	<b>1.00</b>

# Trajectoires d'agression avec une arme en incluant la variable pairs déviants (lagged) au sein des trajectoires

