

*Étude longitudinale  
du développement des enfants  
du Québec (ELDEQ)*

Utilisation des pondérations  
transversales et longitudinales

Institut de la statistique du Québec

Nathalie Plante

Robert Courtemanche

## *Introduction*

- Enquête longitudinale annuelle depuis 1998
- Cohorte de nouveau-nés
- $\approx 2\,800$  familles échantillonnées au volet 1998, dont  $\approx 2\,100$  familles répondantes
- Enfants âgés de 7 ans au volet 2005 (1<sup>ère</sup> année)
- $\approx 1\,500$  répondants au volet 2005
- Objectif principal: connaître les facteurs précurseurs de la réussite scolaire
- Thèmes: développement, santé, alimentation, sommeil, comportement, tempérament, caractéristiques socioéconomiques

## *Population visée*

- **Population visée au volet 1998**
  - enfants nés en 1998 de mères vivant au Québec (naissance simple)
- **Population visée aux volets ultérieurs**
  - enfants nés en 1998 de mères vivant au Québec (naissance simple) et qui n'ont pas quitté la province de façon définitive
- **Pas d'objectif transversal (sauf volet 1998)**
  - pas de renouvellement de l'échantillon pour immigration
  - sous-couverture de l'ordre de 1 % / an  
(en 2004,  $\approx$  8 % des enfants de 1<sup>ère</sup> année étaient nés hors Québec)

## *Plan de sondage et collecte*

- Base de sondage: Fichier maître des naissances vivantes du Québec
- Plan de sondage aléatoire stratifié à 2 degrés:
  1. unité géographique
  2. enfant

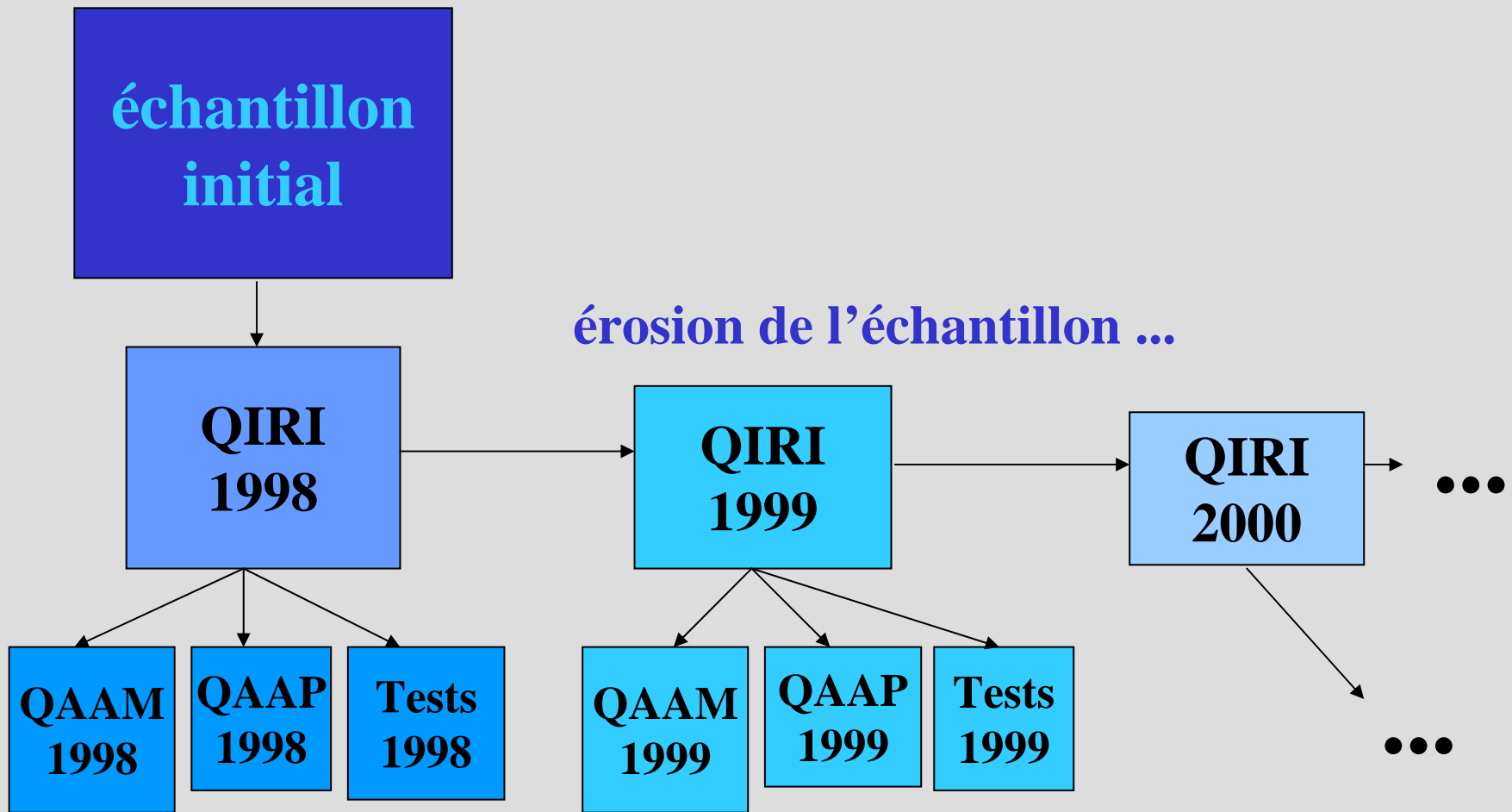
→ léger effet de grappe
- Première collecte à 5 mois, puis annuellement
- Entrevue en face-à-face auprès de la personne qui connaît le mieux l'enfant (sauf volet 2003)

## *Instrument principaux*

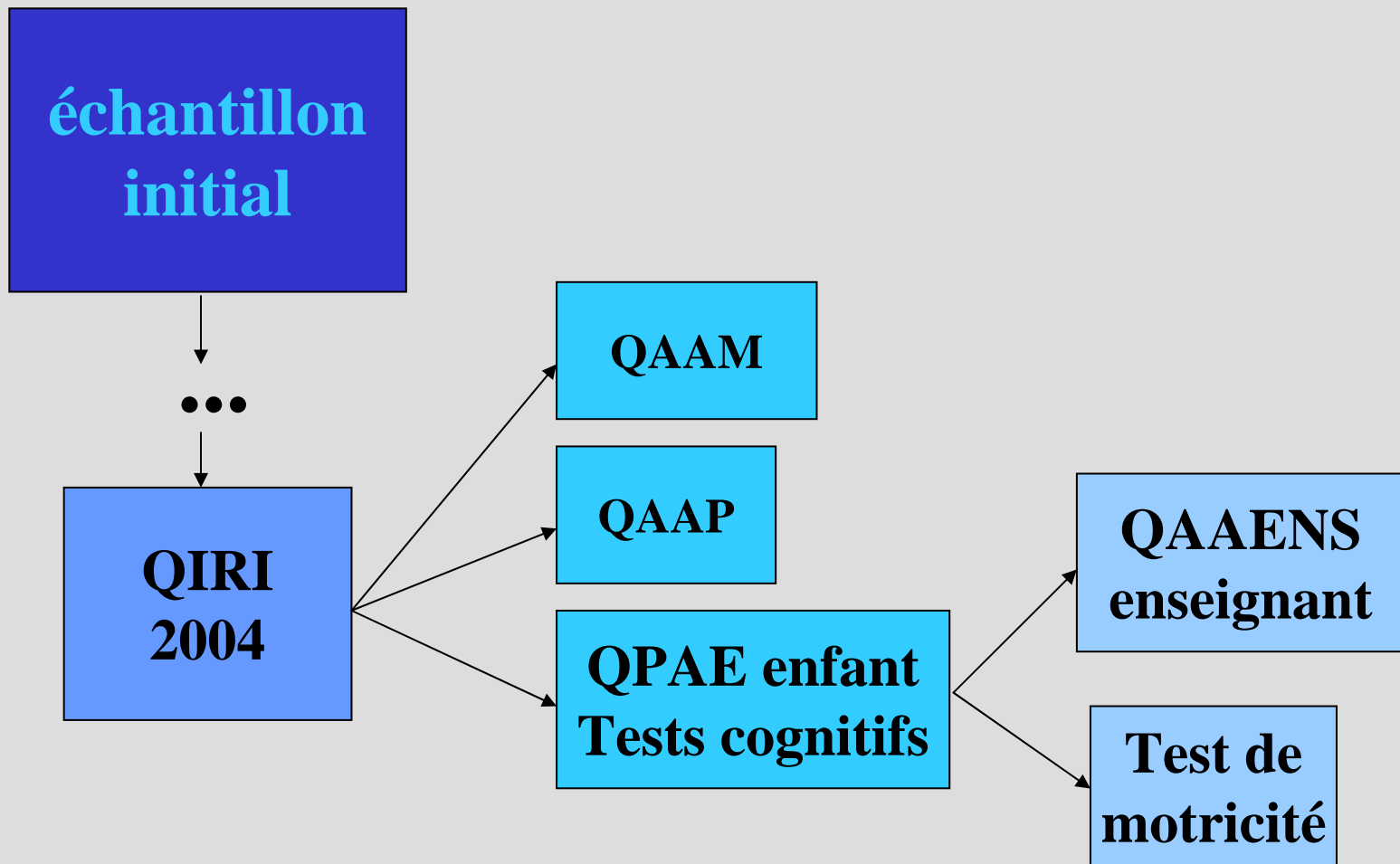
- Questionnaires remplis par l'intervieweuse
  - q. informatisé (**QIRI**) + q. papier (**QPRI**) : 1998 à 2005
- Questionnaire autoadministré de la mère et du père
  - **QAAM** : 1998 à 2005; **QAAP** : 1998 à 2004
  - sur place ou retour par la poste; parents biologiques non cohabitants
- Tests cognitifs administrés à l'enfant
  - varient d'un volet à l'autre de 1998 à 2005
- Test de motricité (2004)
- Questionnaire autoadministré de l'enseignant
  - **QAAENS** : 2004 et 2005
- Questionnaire papier administré à l'enfant
  - **QPAE** : 2004 et 2005



# *Non-réponse hiérarchique*



## *Non-réponse hiérarchique – volet 2004*



## Résultats de collecte

### Questionnaire rempli par l'intervieweur

enquête  
planifiée  
sur 5 ans  
au départ

volet	nombre de répondants	proportion p/r rép. 1998	taux de réponse
1998	2 120		75 %
1999	2 045	96 %	73 %
2000	1 997	94 %	71 %
2001	1 950	92 %	69 %
2002	1 944	92 %	69 %
2003	1 759	83 %	62 %
2004	1 492	70 %	54 %
2005	1 528	72 %	≈ 55 %

## *Représentativité de l'échantillon*

- Du point de vue de la sous-couverture de la population en 1998 ( $\approx 5\%$ )
  - région Nord-du-Québec, territoires cri et inuit et réserves indiennes
  - enfants de l'Outaouais nés en Ontario
  - enfants nés avant la 24<sup>ème</sup> ou après la 42<sup>ème</sup> semaine de grossesse
- Du point de vue de la sous-couverture transversale en 2004
  - $\approx 8\%$  des enfants de 1<sup>ère</sup> année nés hors Québec
  - enfants caractérisés?
- Du point de vue longitudinal
  - Définition de la population visée:  
enfants nés en 1998 de mères vivant au Québec (naissance simple)  
et qui n'ont pas quitté la province de façon définitive

## État de la situation au volet 2004

Langue d'usage de la mère		
n	rép. 1998	rép. 2004
français seulement	1 711	1 250
ni français ni anglais	128	53
autres	279	188

Langue d'usage de la mère			
%	rép. 1998 poids 1998	rép. 2004	
		poids 1998	poids 2004
français seulement	75,7	79,9	75,0
ni français ni anglais	9,2	5,4	8,8
autres	15,1	14,7	16,2

Scolarité de la mère		
n	rép. 1998	rép. 2004
DES ou -	564	355
> DES	1 553	1 136

Scolarité de la mère			
%	rép. 1998 poids 1998	rép. 2004	
		poids 1998	poids 2004
≤ DES	26,6	23,8	27,4
> DES	73,4	76,2	72,6

<b>Niveau de suffisance du revenu à 5 mois</b>		
<b>n</b>	<b>rép. 1998</b>	<b>rép. 2004</b>
suffisant	1 571	1 171
insuffisant	235	154
très insuffisant	276	148

<b>Niveau de suffisance du revenu à 5 mois</b>			
<b>%</b>	<b>rép. 1998 poids 1998</b>	<b>rép. 2004</b>	
		<b>poids 1998</b>	<b>poids 2004</b>
suffisant	72,1	77,2	72,6
insuffisant	12,0	10,9	12,0
très insuffisant	15,9	11,9	15,4

<b>Statut d'immigrant de la mère</b>		
<b>n</b>	<b>rép. 1998</b>	<b>rép. 2004</b>
immigrante non européenne	191	90
autres	1 927	1 401

<b>Statut d'immigrant de la mère</b>			
<b>%</b>	<b>rép. 1998 poids 1998</b>	<b>rép. 2004</b>	
		<b>poids 1998</b>	<b>poids 2004</b>
immigr. non européenne	12,4	8,1	13,6
autres	87,6	91,9	86,4

# *Résultats de collecte*

## Questionnaires autoadministrés des parents

volet	Taux de réponse (proportion de répondants parmi répondants QIRI)	
	QAAM	QAAP
1998	73 % (97 %)	68 % (90 %)
1999	70 % (96 %)	65 % (89 %)
2000	68 % (97 %)	64 % (90 %)
2001	68 % (98 %)	61 % (87 %)
2002	67 % (97 %)	59 % (85 %)
2003	50 % (81 %)	48 % (77 %)
2004	47 % (87 %)	42 % (77 %)
2005	≈ 47 % (87 %)	
<b>1998-2004</b>	<b>39 % (72 %)</b>	<b>≈ 32 % (59 %)</b>



## *Résultats de collecte*

### **Tests cognitifs**

- Tests cognitifs de 1998 à 2005
  - capacité attentionnelle, mémoire, vocabulaire, numération, « school-readiness », lecture, etc.
  - 1998 à 2002: taux de réponse  $\approx$  **60 % à 65 %**
  - 2003: taux de réponse  $\approx$  **35 %**
  - 2004: taux de réponse  $\approx$  **40 %**
  - 2005: taux de réponse  $\approx$  **50 %**

## *Résultats de collecte 2004-2005*

### QPAE, QAAENS et test de motricité

<b>instrument</b>	<b>2004</b>	<b>2005</b>
<b>QPAE enfant</b>	<b>40 %</b>	<b>≈ 50 %</b>
<b>QAAENS enseignant</b>	<b>33 %</b>	<b>≈ 45 %</b>
<b>Test de motricité</b>	<b>36 %</b>	

## *Qu'est-ce qu'une pondération ?*



- Poids = nombre d'enfants qu'un répondant représente dans la population visée
  - ex. : échantillon aléatoire simple de 1 % des enfants de la population (tous répondants):  
poids = 100 pour tous les répondants

## *Pourquoi utiliser une pondération ?*

- Plan de sondage complexe
  - probabilités de sélection variables
- Non-réponse importante
  - hiérarchie → érosion annuelle, non-réponse cumulative
  - à minimiser lors de la collecte
- Non-réponse non aléatoire
  - varie selon caractéristiques des enfants échantillonnés (ex: niveau socioéconomique)
- Inférence à la population visée (non à l'échantillon)
- Minimisation des biais potentiels

## *Estimations pondérées et non pondérées*

<b>Niveau de suffisance du revenu à 5 mois</b>			
<b>%</b>	<b>rép. 1998 poids 1998</b>	<b>rép. 2004</b>	
		<b>poids 2004</b>	<b>aucun poids</b>
suffisant	72,1	72,6	79,5
insuffisant	12,0	12,0	10,5
très insuffisant	15,9	15,4	10,0

<b>Niveau de suffisance du revenu très insuffisant (répondants longitudinaux 1998-2004)</b>							
<b>%</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>
estimations pondérées	14,6	11,1	10,0	6,5		8,4	8,2
estim. non pondérées	9,5	7,2	7,5	5,1		6,0	5,9

## *Estimations pondérées et non pondérées*

Niveau de suffisance du revenu Estimations pondérées			Niveau de suffisance du revenu Estimations non pondérées		
% 1998 \ 2004	revenu suffisant	revenu insuffisant	% 1998 \ 2004	revenu suffisant	revenu insuffisant
revenu suffisant	66,7	5,9	revenu suffisant	73,5	6,1
revenu insuffisant	12,5	14,9	revenu insuffisant	10,3	10,1

# *Pondération transversale et longitudinale*

## Analyse transversale

- Portrait à un temps donné (volet)
- Unités d'analyse: répondants à un volet donné
- Données provenant d'un ou plusieurs instruments
- Poids transversaux (ajustés pour la non-réponse au volet considéré)

## Analyse longitudinale

- Plusieurs volets considérés simultanément
- Liens entre les volets (évolution, cause à effet, etc.)
- Unités d'analyse: répondants à chacun des volets considérés (inter.)
- Poids longitudinaux (ajustés pour la non-réponse à un volet ou l'autre)

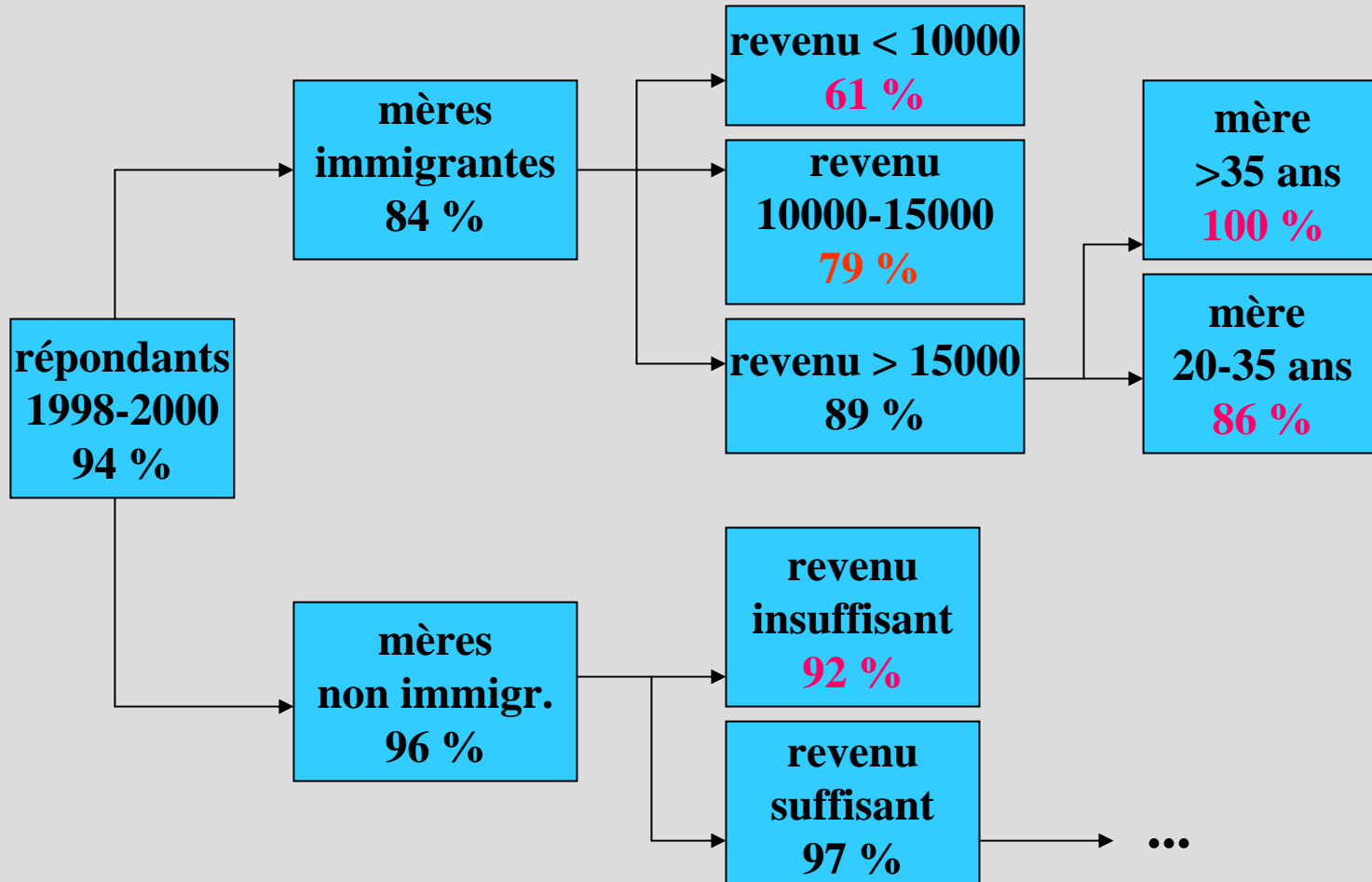
## Poids QIRI

- Hiérarchie de la non-réponse d'un volet à l'autre (approx.)
- Poids transversaux peuvent être considérés comme poids longitudinaux pour analyses longitudinales

## *Comment calcule-t-on les poids ?*

- Poids calculé pour chacun des répondants
- Poids de base = inverse des probabilités de sélection
- Ajustement pour la non-réponse:
  - classes de pondération t.q. proportion de répondants homogène dans une classe
  - variables utilisées liées aux mesures (lorsque possible)
  - ex: classes selon le statut d'immigrant de la mère
  - poids de base divisé par la proportion pondérée de répondants de la classe de pondération
- Poststratification selon région, sexe et durée de grossesse t.q.  $\Sigma$  poids = nb enfants dans popul.

# Classes de pondération (algorithme CHAID)



## *Exemple de calcul des poids*

<b>Enfant</b>	<b>Probabilité de sélection</b>	<b>Poids de base</b>	<b>Classe de pondér.</b>	<b>Taux de réponse</b>	<b>Poids final</b>
<b>1</b>	<b>0,05</b>	<b>20</b>	<b>1</b>	<b>50 %</b>	<b>40</b>
<b>2</b>	<b>0,04</b>	<b>25</b>	<b>1</b>	<b>50 %</b>	<b>50</b>
<b>3</b>	<b>0,10</b>	<b>10</b>	<b>1</b>	<b>50 %</b>	<b>20</b>
<b>4</b>	<b>0,125</b>	<b>8</b>	<b>2</b>	<b>80 %</b>	<b>10</b>
<b>5</b>	<b>0,05</b>	<b>20</b>	<b>2</b>	<b>80 %</b>	<b>25</b>
<b>6</b>	<b>0,10</b>	<b>10</b>	<b>2</b>	<b>80 %</b>	<b>12,5</b>
<b>7</b>	<b>0,04</b>	<b>25</b>	<b>2</b>	<b>80 %</b>	<b>31,25</b>
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>

## *Non-réponse hiérarchique*

- Poids de base: inverse des probabilités de sélection
- Ajustement pour la non-réponse au QIRI
  - ajustement transversal  $\approx$  longitudinal
- Ajustements additionnels pour la non-rép. aux autres instruments
  - transversal ou longitudinal
  - dans certains cas, hiérarchie de la non-réponse entre les différents instruments

## *Poids populationnels et poids échantillonnaux*

- **Soit**  $N =$  taille de la population  
 $n =$  nombre de répondants  
 $\Sigma =$  somme sur les répondants
- $\Sigma$  poids populationnels =  $N$
- **Poids échantillonnal = poids pop.  $\times n/N$**   
t.q.  $\Sigma$  poids échantillonnaux =  $n$   
ou  $\Sigma$  poids échantillonnaux /  $n = 1$
- **Logiciels courants t.q. SAS ou SPSS:**
  - poids échantillonnaux
  - sauf 4 procédures SAS spécifiques pour plan de sondage complexe
  - estimations de précision et tests statistiques approximatifs
- **Fichiers de pondération ELDEQ: poids échantillonnaux**

## *Fichiers de pondération*

- Poids échantillonnaires 1998 à 2004
- Volets 1999 à 2004
  - Fichiers poids201, poids301, ..., poids701:  
poids QIRI, QAAM, QAAP, QPAE, QAAENS  
transversaux et longitudinaux
  - Fichiers poids\_jeux401, ..., poids\_jeux701:  
poids tests cognitifs et de motricité
- Volet 1998
  - Indi101: poids QIRI, QAAP
  - Ipo101: poids IPO transversal

## *Pondérations disponibles*

### *Poids transversaux*

volet	QIRI QPRI	QAAM mère	QAAP père	QPAE enfant	QAAENS enseignant	Tests cognitifs motricité
1998	✓	✓	✓			✓
1999	✓	✓				
2000	✓	✓				
2001	✓	✓				✓
2002	✓	✓				✓
2003	✓	✓	✓			✓
2004	✓	✓	✓	✓	✓	✓

**Poids pour tous les répondants à un instrument et un volet donnés**

✓ : poids spécifique à l'instrument    ✓ : utilisation des poids QIRI

## *Pondérations disponibles*

### *Poids longitudinaux*

volets	QIRI QPRI	QAAM mère	QAAP père	Test cognitif IPO
1998-1999	✓	✓	✓	
1998-2000	✓	✓	✓	
1998-2001	✓	✓	✓	
1998-2002	✓	✓	✓	
1998-2003	✓	✓	✓	
1998-2004	✓	✓		
1999-2001				✓

Poids pour tous les répondants à un instrument et aux volets considérés

✓ : poids spécifique à l'instrument    ✓ : utilisation du poids QIRI

## *Noms des variables de pondération*

exemple: PEQAAML6

1. PE : « poids échantillonnal »
2. Nom de l'**instrument** ou du test cognitif
  - sauf QIRI: PCM – tests regroupés: JEUX
3. L si poids longitudinal (rien autrement)
  - poids QIRI: transversal ou longitudinal
4. Numéro de **volet** (dernier volet si longitudinal)
  - 1 à 7 pour volets 1998 à 2004

Note: Volets 1998 à 2000 → autres dénominations

## *Sélection d'une pondération*

### *Analyse des données - volet 2004 et autres*

<b>Provenance des variables d'analyse</b>	<b>Pondération appropriée</b>
<b>QIRI 2004, seul ou avec QIRI d'autres volets*</b>	<b>PEPCM7</b>
<b>QAAM 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEQAAM7</b>
<b>QAAP 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEQAAP7</b>
<b>Tests cognitifs 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEJEUX7</b>
<b>QPAE 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEQPAE7</b>
<b>QAAENS 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEQAAENS7</b>
<b>Test de motricité, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEMOTRI7</b>
<b>Tests cognitifs et QPAE 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEQPAE7</b>
<b>Tests cognitifs et QAAENS 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEQAAENS7</b>
<b>Tests cognitifs et de motricité 2004, seul ou avec QIRI (un ou pl. volets)</b>	<b>PEMOTRI7</b>
<b>QPAE et QAAENS, seul ou avec tests cognitifs ou avec QIRI</b>	<b>PEQAAENS7</b>

\* 1492 répondants QIRI 2004 vs 1434 répondants 1998 à 2004

## *Critères pour la sélection d'une pondération*

**Lorsque ces situations d'analyse ne s'appliquent pas:**

- ✓ **Minimiser la perte d'unités d'analyse (biais, précision)**
- ✓ **Maximiser la proportion de la non-réponse pour laquelle un ajustement a été fait (biais)**

⇒ **COMPROMIS !!**

- 1. Choisir une pondération**
- 2. Calculer la proportion d'unités d'analyse perdues p.c.q. pas de poids**
- 3. Calculer la proportion des non-répondants parmi les unités pour lesquelles un poids existe**
- 4. Évaluer les pondérations alternatives**
- 5. Choisir la meilleure pondération en mettant l'emphase sur la minimisation de la perte d'unités d'analyse**

## *Choix d'une pondération Exemple 1*

- Analyse longitudinale sur la persistance des troubles de sommeil - volets 1998, 2001 et 2004
- Instrument QAAM 1998, 2001 et 2004
  - avec ou sans QIRI
- Pas de pondération optimale disponible
- Possibilités: PEQAAML7 ou PEQAAM7
- Dans les 2 cas, ajustement pour la non-réponse à partir du poids de référence PEPCM7
  - PEQAAML7: ajustement pour la non-réponse aux volets 1998 à 2004
  - PEQAAM7: ajustement pour la non-réponse au volet 2004 seulement
  - Poids idéal: ajustement pour la non-rép. aux volets 1998, 2001 et 2004

- Unités d'analyse visées (**vert**) :  $n = 1\ 275$
- PEQAAML7 (**gris**) : - perte de 200 unités, soit  $16\ \% = (1\ 275 - 1\ 075) / 1\ 275$ 
  - toute la non-rép. prise en compte (sauf partielle)
- PEQAAM7 (**bleu**) : - aucune perte d'unités d'analyse
  - ne tient pas compte de 3 % de la non-réponse  $(1319 - 1275) / 1319 \rightarrow$  biais?

QIRI 2004:  $n=1\ 492$

PEPCM7

QAAM 2004:  $n=1\ 319$

PEQAAM7

QAAM 1998-2001-2004:  $n=1\ 275$

QAAM 1998 à 2004:  $n=1\ 075$

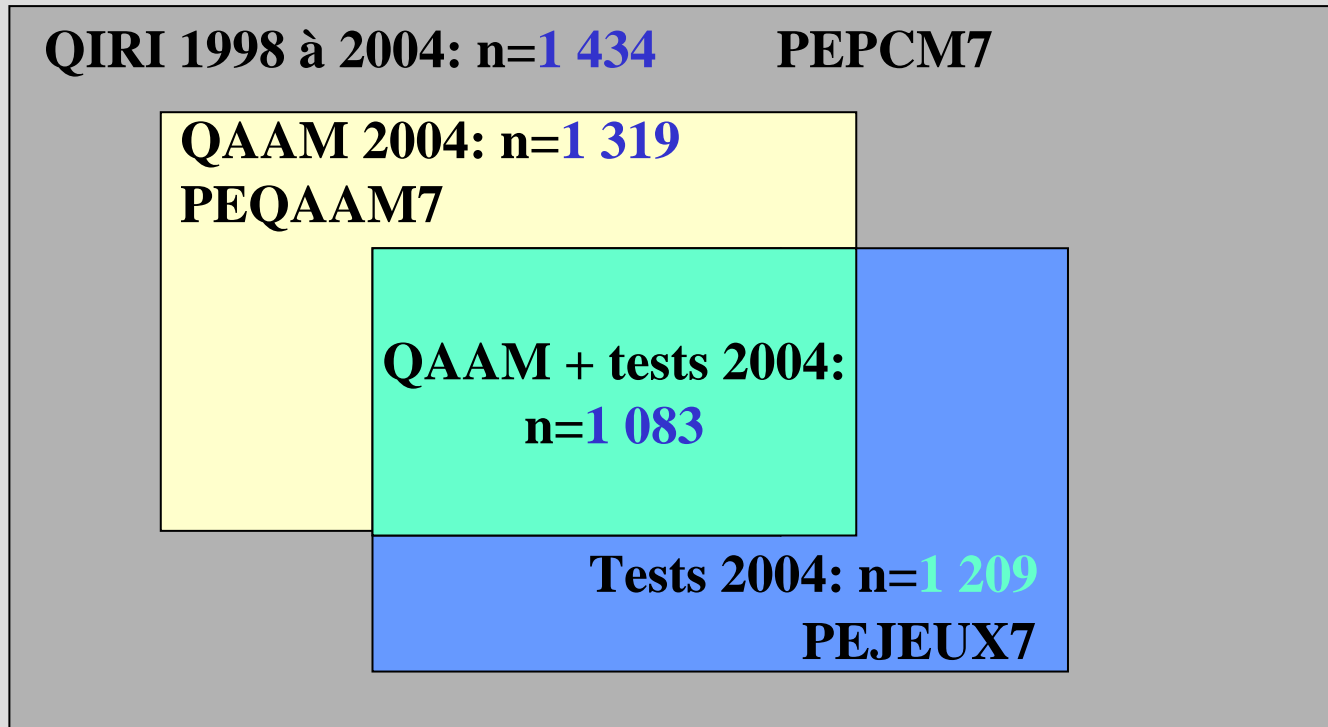
PEQAAML7

## *Choix d'une pondération*

### *Exemple 2*

- Analyse longitudinale QIRI 1998 à 2004 + QAAM et tests cognitifs 2004
- 1 082 unités d'analyse visées
- PEQAAM7 ou PEJEUX7 ?
- Dans les 2 cas:
  - poids de référence PEPCM7
  - aucune perte d'unités d'analyse

- Unités d'analyse visées (**vert**) : n=1 083
- PEQAAM7 (**jaune**) : ne tient pas compte de 18 % de la non-réponse  
 $(1\ 319 - 1\ 083) / 1\ 319 \rightarrow$  biais?
- PEJEUX7 (**bleu**) : ne tient pas compte de 10 % de la non-réponse  
 $(1\ 209 - 1\ 083) / 1\ 209 \rightarrow$  biais?



## *Non-réponse partielle*

- **Non-réponse à une question donnée parmi les répondants à un instrument**
  - $\neq$  non-réponse à un instrument
  - n'est pas prise en compte par la pondération
- **Variables pour lesquelles la non-réponse partielle > 5 % ont déjà été identifiées...**
  - consulter document sur la non-réponse partielle
- **Non-réponse partielle cumulative**
  - combinaison de plusieurs variables provenant d'instruments et de volets différents
- **Évaluer la perte d'unités d'analyse due à la non-réponse partielle**
- **Évaluer le sens du biais potentiel**
  - étudier les caractéristiques des non-répondants

## *Exemple de non-réponse partielle*

Évaluation de l'enfant par l'enseignant quant  
aux capacités liées à la lecture  
(volet 2004; fin maternelle)

- Relier des sons à des lettres, participer à des activités de lecture en groupe (tâches « faciles »)
  - + de non-réponse partielle chez enfants moins bons aux tests cognitifs
  - surestimation de la prop. d'enfants capables de réaliser ces tâches ?
- Lire des mots complexes, des phrases simples (tâches « difficiles »)
  - + de non-réponse partielle chez enfants meilleurs aux tests cognitifs
  - sous-estimation de la prop. d'enfants capables de réaliser ces tâches ?

## *Estimation*

- Doit tenir compte du plan de sondage complexe:
  - estimation ponctuelle (proportion, moyenne, paramètre d'un modèle, etc.)
  - estimation de la variance
  - statistique d'un test et seuil observé
- Idéalement: utilisation d'un logiciel permettant de tenir compte du plan de sondage

## *Estimation à l'aide d'un logiciel ne permettant pas de tenir compte du plan de sondage*

- **Ces logiciels peuvent en général tenir compte de la pondération**
- **Avec poids populationnels:**
  - estimations ponctuelles  $\approx$  sans biais
  - variances sous-estimées de façon très importante
  - rejet beaucoup trop fréquent de l'hypothèse  $H_0$
- **Avec poids échantillonnaux:**
  - estimations ponctuelles  $\approx$  sans biais
  - si les probabilités d'être répondant sont variables:
    - variances en général sous-estimées
    - rejet trop fréquent de l'hypothèse  $H_0$

## *Solutions*

- **Situation**
  - pas de logiciel tenant compte du plan de sondage
  - pas de données individuelles sur le plan de sondage
  - modèle non supporté par de tels logiciels
- **Estim. d'une proportion et test approx. du  $\chi^2$  :**
  - ajustement de la variance et de la statistique du test à l'aide d'un effet de plan moyen (résultat approx.):
    - diviser les poids divisés par l'effet de plan moyen
    - multiplier la variance par l'effet de plan moyen (tous estim.)
- **Autres tests ou modèles :**
  - abaissement du seuil (ex: 0,01 au lieu de 0,05)
    - approche conservatrice pour l'ELDEQ
    - en tenir compte dans l'interprétation

## *Effet de plan*

- Mesure de l'efficacité du plan de sondage
- Effet de plan = rapport de la variance obtenue sur la variance qui aurait été obtenue avec un plan aléatoire simple de même taille
- ELDEQ: estimation de l'effet de plan moyen des proportions estimées à partir des données recueillies
  - effet de plan moyen pour une proportion = **1,3**
  - plan de sondage de l'ELDEQ à 2 degrés: moins efficace que E.A.S. mais entrevue moins coûteuse!

*Des questions ???*