



**Étude longitudinale du développement des enfants du Québec (ÉLDEQ):  
pourquoi et comment considérer  
le plan de sondage complexe dans les analyses**

**Luc Belleau, Robert Courtemanche et Hélène Desrosiers**

**Institut de la statistique du Québec**

**Présentation au Centre interuniversitaire québécois de statistiques sociales**

**Montréal, 22 avril 2010**



## *Plan de la présentation*

- 1) Présentation de l'ÉLDEQ : description et rappel de certains concepts
  
- 2) La pondération et le plan de sondage complexe
  - Définition et pourquoi en tenir compte
  - Illustrations
  
- 3) Comment tenir compte du plan de sondage
  - Mises en situation & outils disponibles

# 1) Présentation de l'ÉLDEQ



## *Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2010)*

- **Particularité** : Cohorte de nouveau-nés
  
- **Objectifs** :
  - Identifier les facteurs de la petite enfance qui contribuent à l'adaptation sociale et à la réussite scolaire des enfants québécois
  
  - Mieux cerner le rôle de programmes publics (ex. : services de garde, programmes de prévention en santé) et orienter le développement de programmes futurs

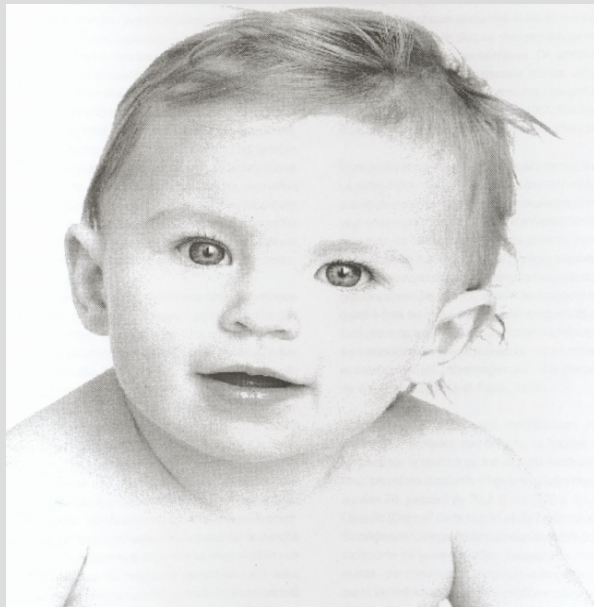


## *L'ÉLDEQ 1998-2010: deux grandes phases de réalisation*

---

- Phase 1 (1998-2002)
  - Période de la petite enfance (collectes E1 à E5)
  
- Phase 2 (2003-2010)
  - Périodes préscolaire et primaire (collectes E6 à E13)

## *Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2010)*



Échantillon initial (1998) = 2 817  
enfants

Répondants initiaux (1998) = 2 120  
enfants âgés d'environ 5 mois

## *Population visée*

- **Population visée au volet 1998**
  - Enfants nés en 1997-1998 de mères vivant au Québec (naissances simples)
- **Population visée aux volets ultérieurs**
  - Enfants nés en 1997-1998 de mères vivant au Québec (naissances simples) 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 2008,  $\approx$  10 % des enfants de 4<sup>e</sup> année étaient nés hors Québec)

## *Plan de sondage*

- Base de sondage: Fichier maître des naissances vivantes 1997-1998 du Québec
- Plan de sondage aléatoire stratifié à 2 degrés :
  1. Unité géographique
  2. Enfant

**→ Léger effet de grappe**



## *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>e</sup> ou après la 42<sup>e</sup> semaine de grossesse
- Du point de vue de la sous-couverture transversale en 2008
  - $\approx 10\%$  des enfants de 4<sup>e</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 (naissances simples) et qui n'ont pas quitté la province de façon définitive

## *La collecte en bref*

### Portrait détaillé des premières années de vie des enfants

➔ Suivi annuel de 5 mois jusqu'à 8 ans (2<sup>e</sup> année),  
puis suivi bisannuel jusqu'à l'âge de 12 ans

Mesures directes (ex. : tests cognitifs, mesures anthropométriques) et indirectes

Différentes sources : parents, enfants, enseignants, intervieweurs, données administratives

## *Le suivi longitudinal des enfants de l'ÉLDEQ*

	<b>Année d'enquête</b>	<b>Âge</b>	<b>Répondants</b>	<b>Coll. spec.</b>
<b>Phase I</b>	1998 (E1)	5 mois	2 120	
	1999 (E2)	17 mois	2 045	
	2000 (E3)	29 mois	1 997	
	2001 (E4)	41 mois	1 950	
	2002 (E5)	44 à 56 mois <sup>1</sup>	1 944	Nutrition : MSSS
<b>Phase II</b>	2003 (E6)	56 à 68 mois	1 759	Milieus de garde : GRIP
	2004 (E7)	68 à 80 mois	1 492	Motricité/cond. phys : QEF
	2005 (E8)	80 à 92 mois	1 528	
	2006 (E9)	92 à 104 mois	1 526	Motricité/cond. phys : QEF
	2008 (E11)	116 à 129 mois	1 402	Volet santé : U de Mtl
	2010 (E13)	≈ 12 ans	?	

<sup>1</sup> À partir de E5, la période de collecte a été modifiée afin de s'adapter à la réalité du système scolaire

# Contenu thématique

## ENFANT

- Données sociodémographiques
- Santé & condition phys.
- Habitudes de vie
- Tempérament
- Compétences sociales
- Maturité affective
- Langage et aptitudes cognitives
- Activités éducatives et littératie
- Relations avec les enseignants
- Motivation scolaire
- Engagement scolaire
- Performance scolaire, etc.



## ENVIRONNEMENT FAMILIAL

- Données sociodémographiques
- Conditions de vie matérielles
- Santé et habitudes de vie
- Antécédents
- Littératie
- Pression du temps
- Capital social
- Relation parents-enfant
- Fonctionnement familial
- Transitions familiales et garde légale, etc.



## ENVIRONNEMENT PHYSIQUE & SOCIAL

- Utilisation des services (garde, santé, aide aux devoirs)
- Lien famille-école
  - Climat scolaire
  - Soutien social aux parents
  - Caractéristiques du quartier, etc



Grossesse - Naissance (1997-1998)

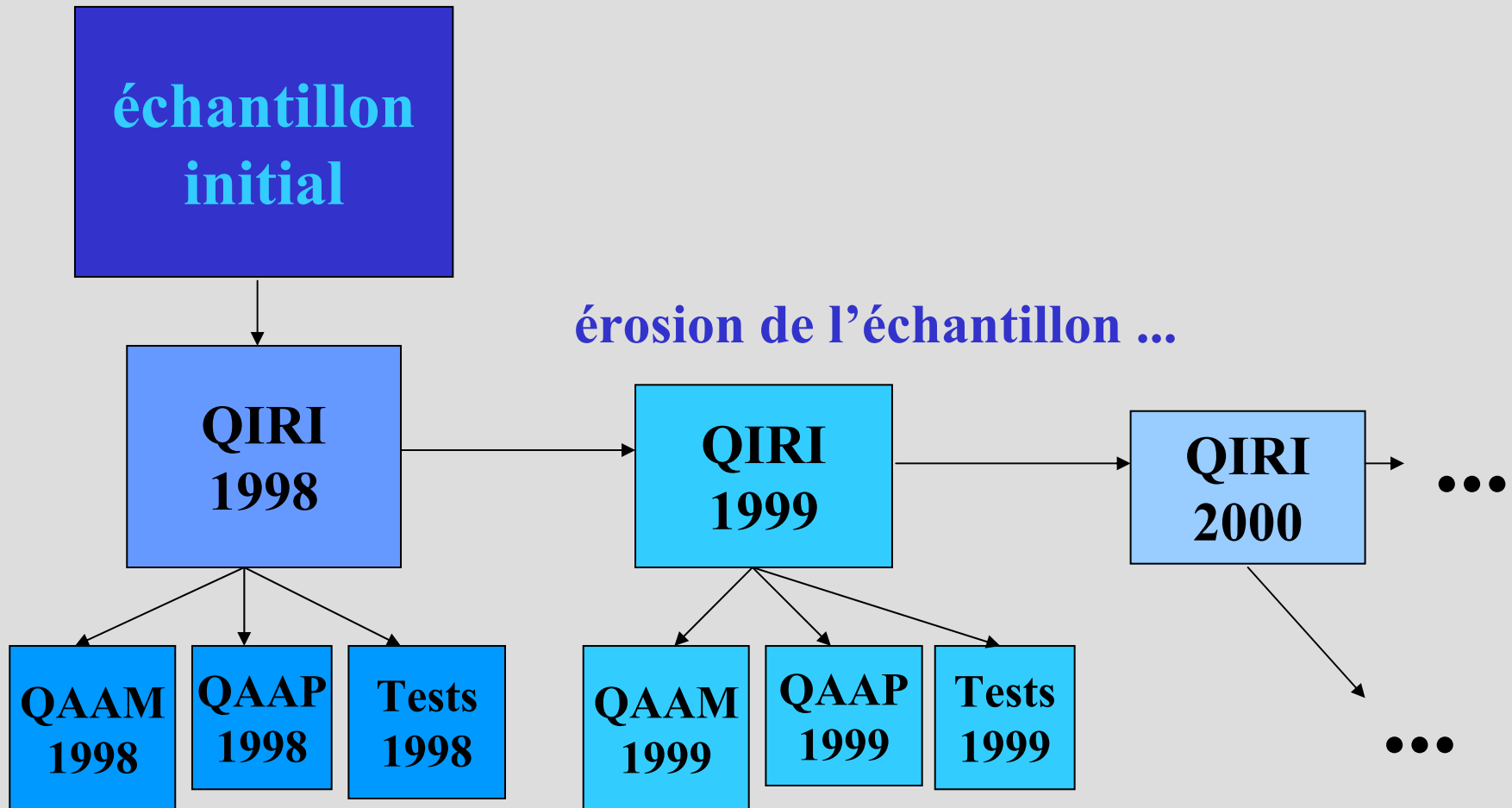
12 ans (2010)...



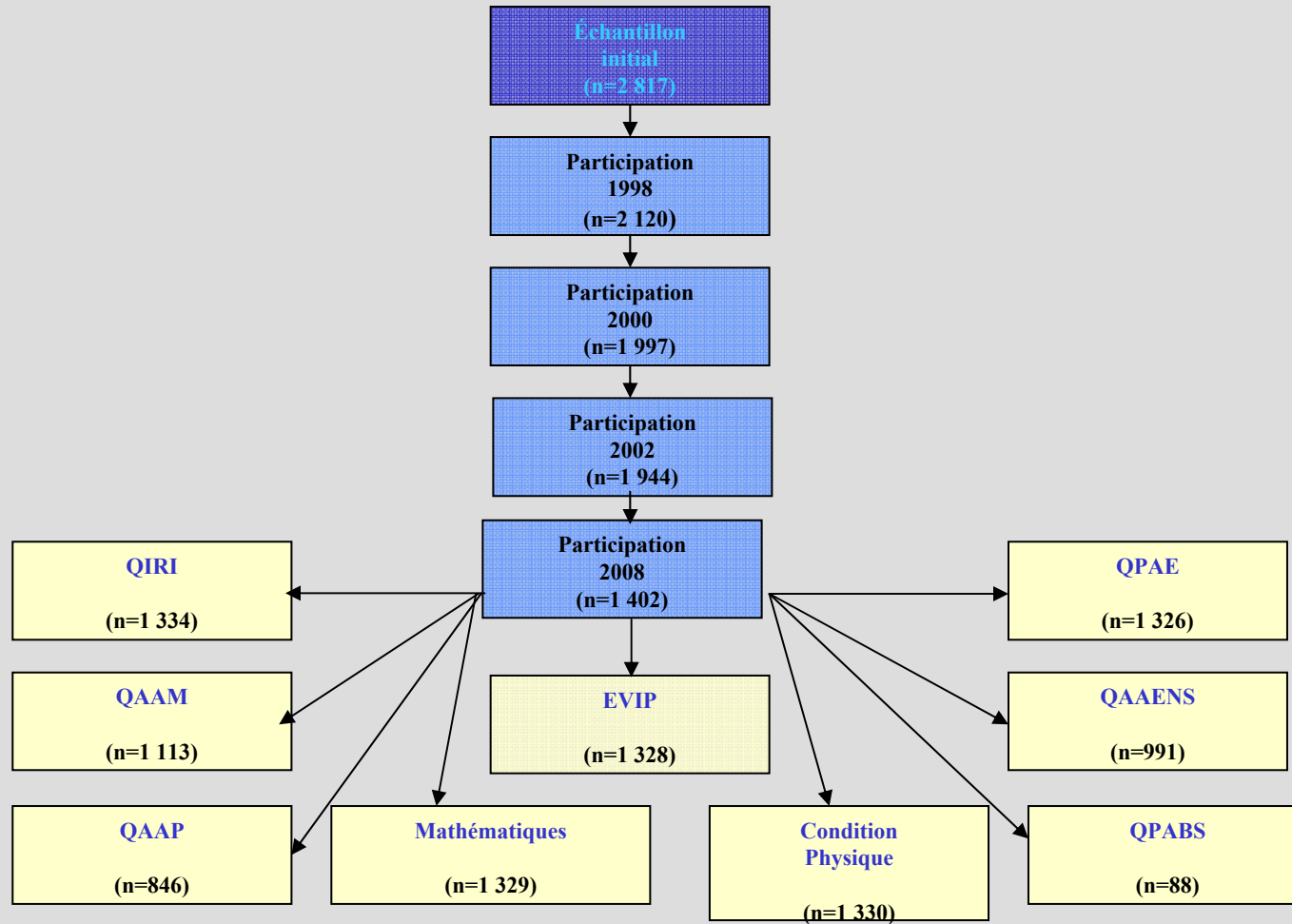
## *Principaux instruments de collecte 1998-2008*

- **Questionnaires remplis par l'intervieweur**
  - q. informatisé (**QIRI**) : 1998 à 2006, 2008
  - q. papier (**QPRI**) : 1998 à 2003, 2006
  - Inventaire du milieu familial (**IMF**) : 1998 à 2000
  - Questionnaire complété par l'intervieweur (**QCI**) : 2001, 2002, 2004 à 2006 et 2008
- **Questionnaires autoadministrés de la mère et du père**
  - **QAAM** : 1998 à 2006, 2008; **QAAP** : 1998 à 2004, 2008
- **Tests cognitifs administrés à l'enfant**
  - Varient d'un volet à l'autre de 1998 à 2006, 2008
- **Tests de motricité et de condition physique (2004 et 2006)**
- **Questionnaire papier administré à l'enfant**
  - **QPAE** : 2004 à 2006, 2008
- **Questionnaire autoadministré de l'enseignant**
  - **QAAENS** : 2004 à 2006, 2008
- **Questionnaire autoadministré de la fratrie**
  - **QAAF** : 2001 et 2003

# *Non-réponse hiérarchique*



## *Hiérarchie des répondants aux différents instruments de collecte*



## Résultats de collecte

### Questionnaire rempli par l'intervieweur

	Volet	Nombre de répondants	Proportion p/r rép. 1998	Taux de réponse
Enquête planifiée sur 5 ans au départ	1998	2 120		75 %
	1999	2 045	96 %	73 %
	2000	1 997	94 %	71 %
	2001	1 950	92 %	69 %
	2002	1 944	92 %	69 %
Maternelle à 4 <sup>e</sup> année	2003	1 759	83 %	62 %
	2004	1 492	70 %	53 %
	2005	1 528	72 %	54 %
	2006	1 526	72 %	54 %
	2008	1 402	66 %	50 %



## Résultats de collecte

### Ex. # 1 : Questionnaires autoadministrés des parents

Volet	Taux de réponse	
	QAAM	QAAP
1998	73 %	68 %
1999	70 %	65 %
2000	68 %	64 %
2001	68 %	61 %
2002	67 %	59 %
2003	50 %	48 %
2004	47 %	42 %
2005	≈ 47 %	
1998-2004	39 %	≈ 32 %

Maternelle  
et 1<sup>ère</sup> année

## *Résultats de collecte*

### Ex. # 2 : Tests cognitifs

- Tests cognitifs de 1998 à 2005
  - Capacité attentionnelle, mémoire, vocabulaire, numératie, préparation à l'école, lecture, etc.
  - 1998 à 2002 : taux de réponse  $\approx$  **60 % à 65 %**
  - 2003 : taux de réponse  $\approx$  **35 %**
  - 2004 (E7, maternelle) : taux de réponse  $\approx$  **40 %**
  - 2005 (E8, 1<sup>e</sup> année) : taux de réponse  $\approx$  **50 %**

# Résultats de collecte

Ex. # 3 : Questionnaire papier administré à l'enfant

Volet	Taux de réponse
2004	42 %
2005	53 %
2006	52 %
2008	47 %
2004-2005	39 %
2004-2006	38 %
2004-2008	34 %

Maternelle  
à 4<sup>e</sup> année →

## Résultats de collecte

Ex. # 4 : Questionnaire autoadministré de l'enseignant

Volet	Taux de réponse
2004	34 %
2005	47 %
2006	46 %
2008	35 %
2004-2005	29 %
2004-2006	26 %
2004-2008	19 %

Maternelle →  
à 4<sup>e</sup> année

## *Les résultats de collecte en bref*

- Le taux de réponse diminue avec le temps
  - Érosion annuelle de l'échantillon
- Les taux de réponse diffèrent d'un instrument à un autre
- Les taux de réponse « *transversaux* » sont meilleurs que les taux de réponse « *longitudinaux* »
  - Répondant transversal (à un volet donné)
  - Répondant longitudinal (répondant à tous les volets considérés)

## 2) La pondération et le plan de sondage complexe

Définition et pourquoi en tenir compte

## *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

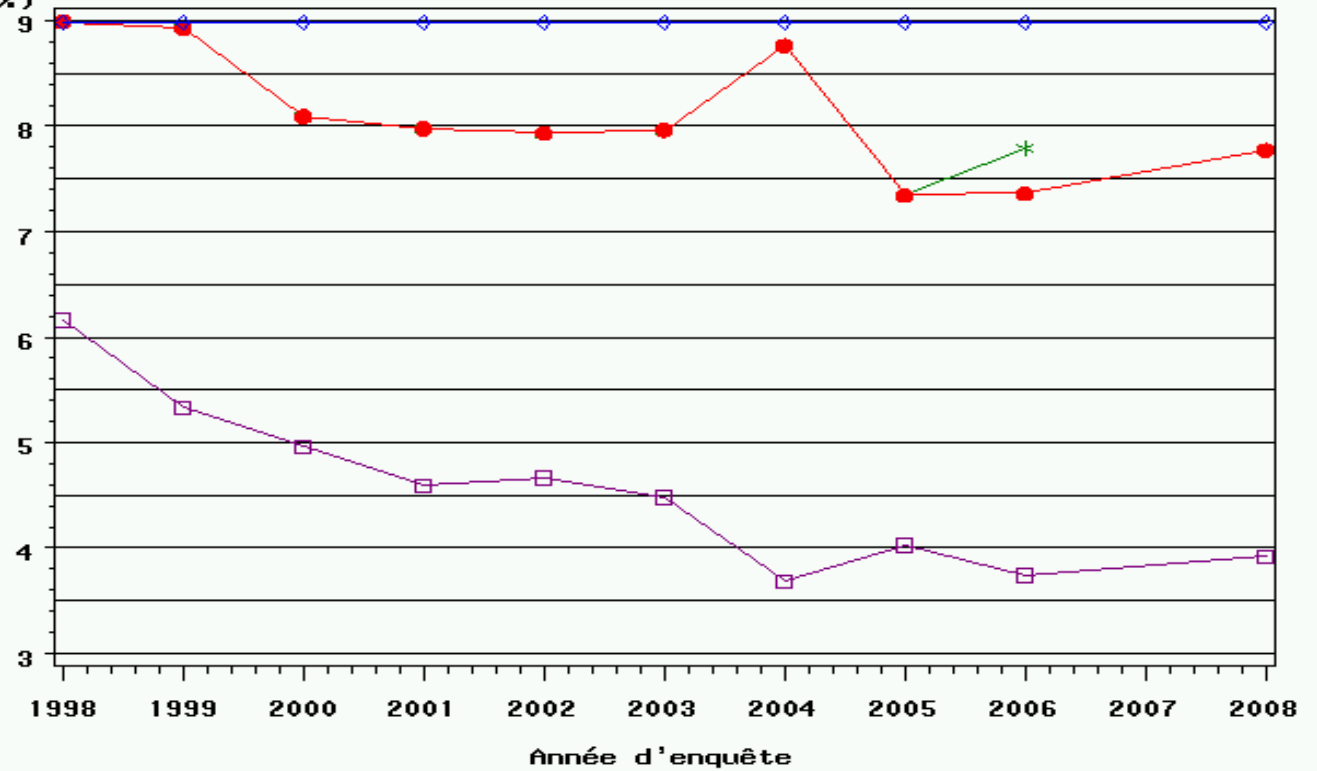
## *Pourquoi utiliser une pondération ?*

- 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)
- Réduction des biais potentiels



Proportion estimée d'enfants dont la mère ne parle ni français, ni anglais en 1998

Proportion d'enfants (en %)

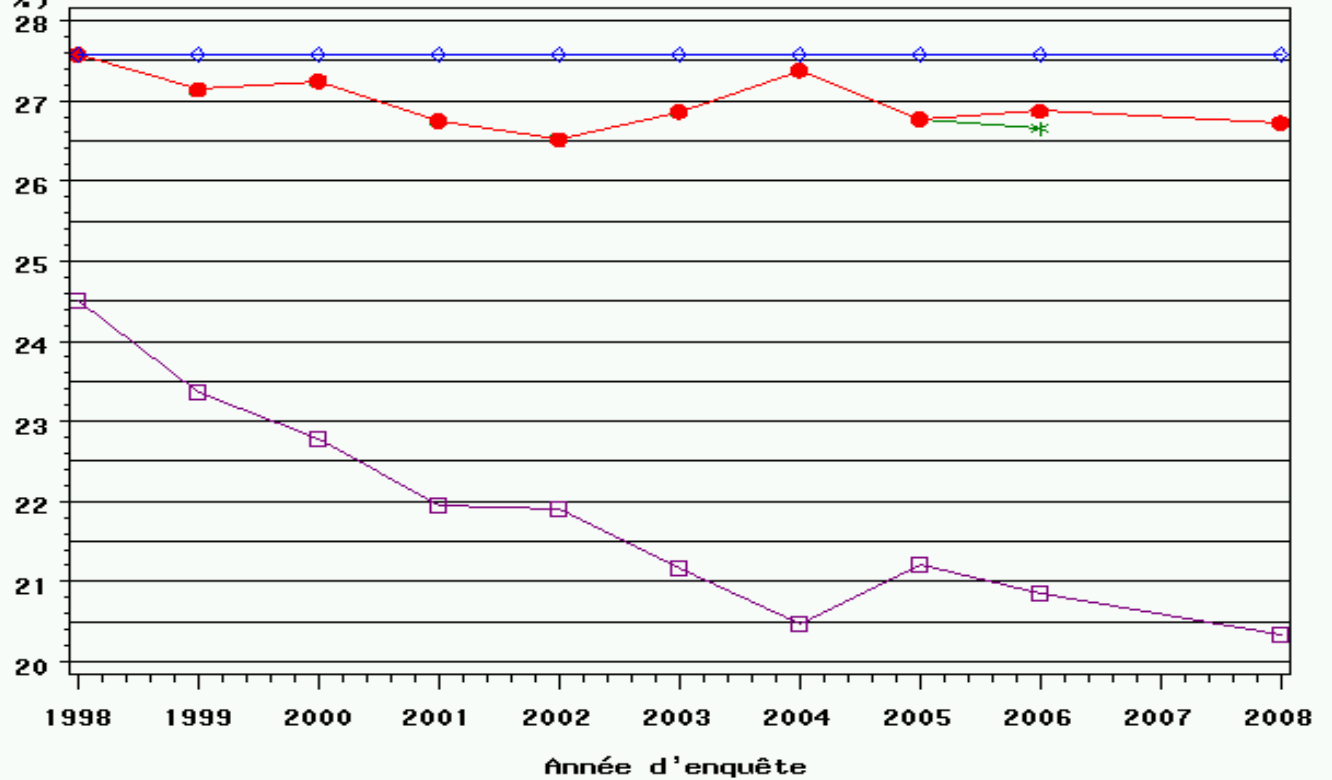


Courbes

- \*-\*-\* Estim. transv. (avec corr. pour la non-réponse): SCORE 2006
- Estim. transv. (avec corr. pour la non-réponse): CHAID 1998-2006 et SCORE 2008
- ◇-◇-◇ Estimation transversale de référence (année 1998)
- Estimations initiales (sans correction pour la non-réponse)

Proportion estimée d'enfants provenant de ménages ayant des revenus insuffisants ou très insuffisants en 1998

Proportion d'enfants (en %)



Courbes

- \*-\*-\* Estim. transv. (avec corr. pour la non-réponse): SCORE 2006
- Estim. transv. (avec corr. pour la non-réponse): CHAID 1998-2006 et SCORE 2008
- ◇-◇-◇ Estimation transversale de référence (année 1998)
- Estimations initiales (sans correction pour la non-réponse)

## *Caractéristiques des non-répondants*

- Faible revenu
- Langue parlée autre que le français et l'anglais
- Famille monoparentale
- Scolarité de la mère : DES et moins
- Âge de la mère en 1998 < 25 ans

# Exemple #1: Fréquence

Type de famille et rendement scolaire global en 4<sup>e</sup> année (E11)

	Au-dessus de la moyenne	
	Non pondéré	Pondéré
Fam. intacte	50,7	48,6
Fam. recomposée	29,2	25,4
Fam. monoparentale	37,1	32,3

## *Pourquoi utiliser une pondération ?*

- Les estimations non pondérées **diffèrent** des estimations pondérées
  - On NE peut PAS faire l'hypothèse que les répondants sont représentatifs de la population visée
  
- Les corrections pour la non-réponse sont **nécessaires** pour que les estimations s'approchent de la « cible »
  - Les ajustements sont assez efficaces, en particulier pour les mesures associées aux caractéristiques de non-répondants

## *Pourquoi tenir compte du plan de sondage ?*

- **Estimations ponctuelles pondérées  $\approx$  sans biais**
- **Si les probabilités d'être répondant sont variables et concentration de l'échantillon :**
  - Variances en général sous-estimées
  - Rejet trop fréquent de l'hypothèse  $H_0$
- **Il faut utiliser un logiciel permettant de tenir compte du plan de sondage**

## *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
- ÉLDEQ : 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'ÉLDEQ à 2 degrés : moins efficient que E.A.S. mais entrevue moins coûteuse!

# Exemple #2: Régression logistique

Surplus de poids chez les enfants de 7 ans (E8)

	Sans pondération ni plan de sondage		Pondérations		Pondérations et plan de sondage (Sudaan) <sup>1</sup>	
	RC et IC	P-value	RC et IC	P-value	RC et IC	P-value
Sexe (fille)	1,29 (0,91;1,83)	<b>0,1513</b>	1,34 (0,96;1,93)	<b>0,0873</b>	1,36 (0,95; 1,95)	<b>0,0962</b>
Durée du sommeil nocturne (10 h ou moins)	1,46 (1,03;2,08)	<b>0,0341</b>	1,43 (0,99;2,06)	<b>0,0536</b>	1,43 (0,93; 2,19)	<b>0,0988</b>
Fréquence de consom. de boissons sucrées ou gazeuses (dernière semaine) <i>(réf. : aucune fois)</i>		<b>0,1322</b>		<b>0,014</b>		<b>0,0188</b>
1 à 2 fois	1,46 (1,0; 2,12)		1,77 (1,20;2,60)		1,77 (1,19;2,62)	
3 fois ou plus	1,13 (0,65;1,95)		1,25 (0,73;2,13)		1,25 (0,71;2,18)	

1. Tiré de Desrosiers, Dumitru et Dubois (2009). Après ajustement pour les variables suivantes : tabagisme durant la grossesse, nombre de parents avec embonpoint, enfant mange trop ou trop vite, statut d'immigrante de la mère et région de résidence.





# Exemple #3: Régression logistique

## Retard de vocabulaire chez les enfants de maternelle (E7)

	Sans pondération ni plan de sondage		Pondérations		Pondérations et plan de sondage (Sudaan) <sup>1</sup>	
	RC et IC	P-value	RC et IC	P-value	RC et IC	P-value
Bébé de petit poids	1,86 (0,74;4,69)	<b>0,1908</b>	2,95 (1,35;6,43)	<b>0,0066</b>	2,95 (1,26;6,92)	<b>0,0131</b>
Infection(s) aux oreilles à 17 mois	1,41 (0,94;2,11)	<b>0,0954</b>	1,6 (1,07;2,38)	<b>0,0215</b>	1,6 (1,02;2,50)	<b>0,0406</b>
Dynamique de faible revenu ( <i>réf. : aucun épisode ou épis. récent seul.</i> )		<b>0,5693</b>		<b>0,0634</b>		<b>0,1283</b>
Avant 4 ans	1,26 (0,80;1,97)		1,61 (1,05;2,50)		1,61 (0,98;2,65)	
Chronique	1,26 (0,60;2,61)		0,99 (0,53;1,87)		0,99 (0,44;2,24)	
Santé < très bonne ( <i>réf. : jamais</i> )		<b>0,2593</b>		<b>0,0192</b>		<b>0,0829</b>
Avant 4 ans	1,43 (0,92;2,21)		1,82 (1,20;2,77)		1,82 (1,08;3,08)	
À 5 ou 6 ans seul.	1,25 (0,69;2,24)		1,21 (0,68;2,17)		1,21 (0,65;2,26)	

1. Tiré de Desrosiers et Ducharme (2008). Après ajustement pour les variables suivantes : sexe, âge, exposition à une autre langue, scolarité de la mère, surprotection maternelle (5 mois), lecture à l'enfant (17 mois) et soutien social (29 mois).

# Exemple #4: Régression linéaire

Rendement scolaire global en première année (E8)

	Sans pondération ni plan de sondage			Pondérations			Pondérations et plan de sondage (Sudaan) <sup>1</sup>		
	Beta	SE	<i>P-Value</i>	Beta	SE	<i>P-Value</i>	Beta	SE	<i>P-Value</i>
Âge de l'enfant	0,01	0,03	0,7656	0,01	0,03	0,6753	0,01	0,03	0,7210
Sexe de l'enfant	0,04	0,03	<b>0,1784</b>	-0,07	0,03	<b>0,0103</b>	-0,07	0,03	<b>0,0219</b>
<b>Jeux du Lollipop</b>									
Jeu 1 – Couleurs, formes	0,12	0,03	0,0000	0,13	0,03	0,0000	0,13	0,13	0,0000
Jeu 2- Repérage spatial	0,09	0,03	0,0009	0,06	0,03	0,0143	0,06	0,03	0,0344
Jeu 2- Chiffres et calcul	0,14	0,03	0,0000	0,14	0,03	0,0000	0,14	0,03	0,0001
Jeu 4- Lettres et écriture	0,44	0,03	0,0000	0,46	0,03	0,0000	0,46	0,03	0,0000

## *Données administratives du MELS : objectif et méthode*

- Vérifier si les estimations de l'ÉLDEQ sont biaisées
- Comparer des proportions calculées à partir de la population visée de l'ÉLDEQ et à partir de l'échantillon de l'ÉLDEQ selon différentes pondérations

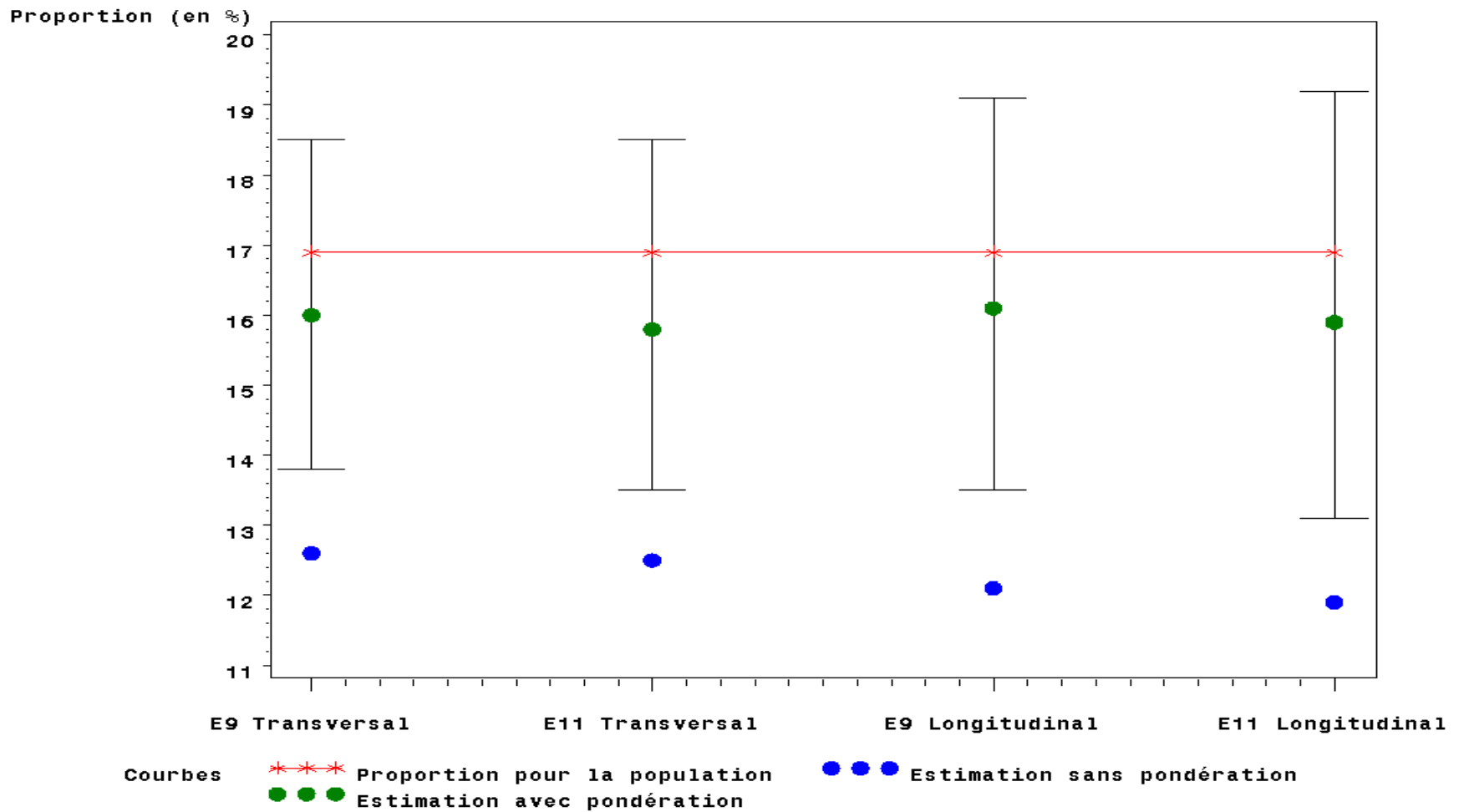
## *Données administratives du MELS : fichier de la population visée*

- Les variables ont été compilées pour tous les enfants nés entre le 1<sup>er</sup> octobre 1997 et le 30 septembre 1998 (inclusivement), qui fréquentaient une école au Québec durant l'année scolaire 2008-2009, en excluant ceux dont le lieu de naissance n'est pas la province de Québec

## *Données administratives du MELS : variables disponibles*

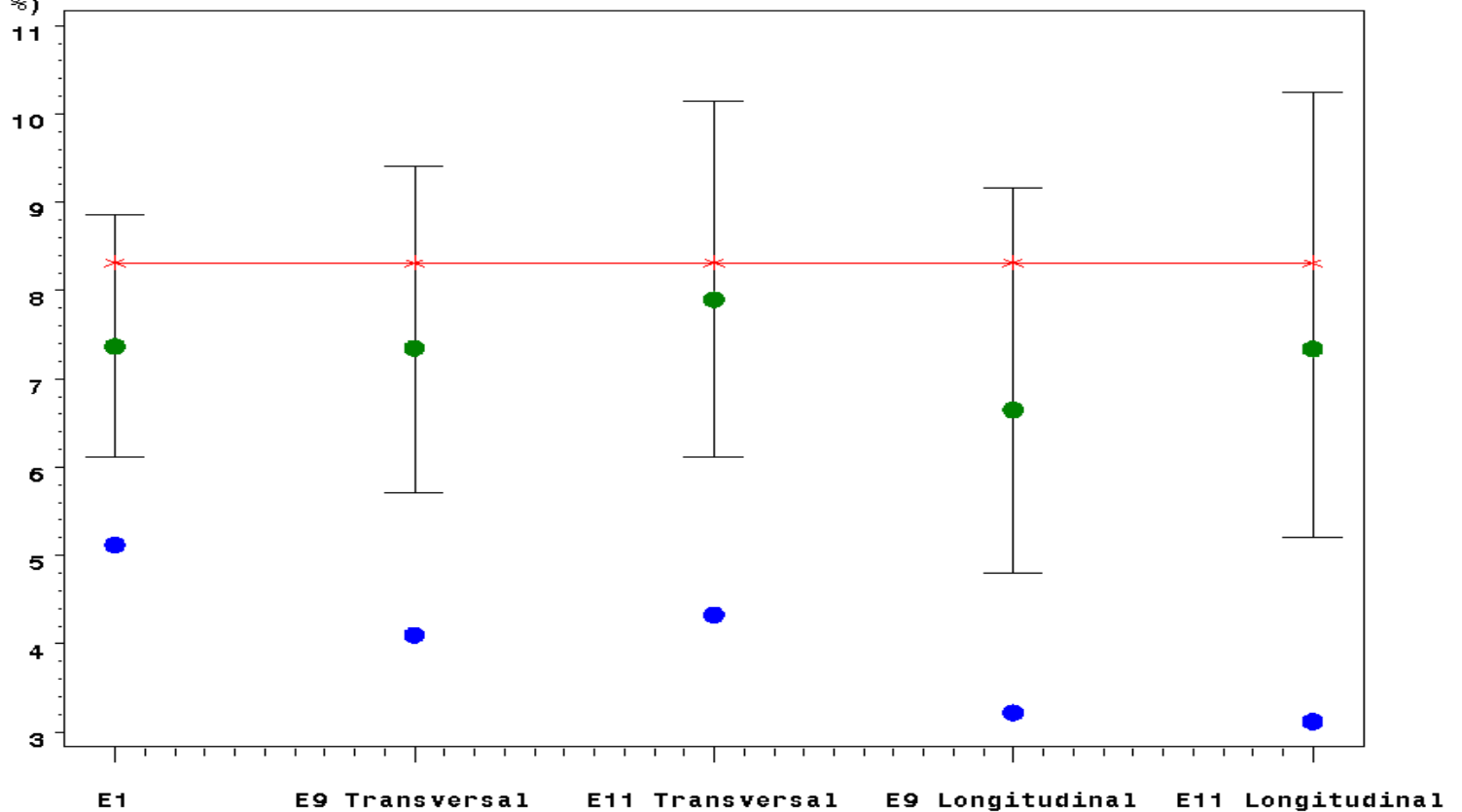
- L'indice du seuil de faible revenu (compilations par décile)
- La langue maternelle de l'enfant

Enfant avec un indice de seuil de revenu à 9 ou 10



Enfant dont la langue maternelle est ni l'anglais, ni le français

Proportion (en %)



Courbes \* \* \* \* Proportion pour la population      • • • Estimation sans pondération  
 • • • Estimation avec pondération

## *Pourquoi pondérer et tenir compte du plan de sondage?*

- En ne pondérant pas ...
  - Les estimations de l'ÉLDEQ diffèrent des vraies valeurs (notamment à cause de BIAIS dû à la non-réponse totale)
- En pondérant ...
  - Les estimations de l'ÉLDEQ s'approchent de la cible (RÉDUCTION DES BIAIS)



## *Pourquoi pondérer et tenir compte du plan de sondage?*

- En pondérant
  - Les estimations ponctuelles sont changées
  
- En tenant compte du plan de sondage ...
  - Les estimations de variance changent
  - Les intervalles de confiance changent
  - **Les conclusions des tests changent!**

### 3) Comment tenir compte du plan de sondage dans l'ÉLDEQ

## *Comment pondérer?*

- Pondération hiérarchique
- Poids de base : inverse des probabilités de sélection
- Ajustement pour la non-réponse pour la participation à chaque volet
- Ajustements additionnels pour la non-réponse aux autres instruments
  - Dans certains cas, hiérarchie de la non-réponse entre les différents instruments

## *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)
- Attention inférence : enfants nés en 1998

### 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)
- Parfois, des poids transversaux peuvent être considérés comme poids longitudinaux pour analyses longitudinales

## *Poids populationnels et poids échantillonnax*

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

## *Fichiers de pondération*

- Poids échantillonnaires 1998 à 2008
- Volets 1999 à 2008
  - Fichiers poids201, poids301, ..., poids1101:  
poids QIRI, QAAM, QAAP  
transversaux et longitudinaux
  - Fichiers poids\_jeux401, ..., poids\_jeux601:  
poids tests cognitifs et de motricité
  - Fichiers poids711, ..., poids1111:  
poids tests cognitifs et de motricité, QPAE, QAAENS
- Volet 1998
  - Indi101: poids QIRI, QAAP
  - Ipo101: poids IPO transversal

## *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  
**T** ou rien si transversal
  - Poids QIRI : transversal ou longitudinal
4. Numéro de **volet** (dernier volet si longitudinal)
  - 1 à 11 pour volets 1998 à 2008

# *Pondérations disponibles*

## *Poids transversaux*

volet	QIRI	QAAM mère	QAAP père	QPAE enfant	QAAENS enseignant	Jeux
1998	✓	✓	✓			✓
1999	✓	✓	✓			
2000	✓	✓	✓			
2001	✓	✓	✓			✓
2002	✓	✓	✓			✓
2003	✓	✓	✓			✓
2004	✓	✓	✓	✓	✓	✓
2005	✓	✓		✓	✓	✓
2006	✓	✓			✓	
2008	✓				✓	



## *Pondérations disponibles* *Poids longitudinaux<sup>1</sup>*

volets	QIRI	QAAM mère	QAAP père	QPAE enfant	QAAENS enseignant
1998-1999	✓	✓	✓		
1998-2000	✓	✓	✓		
1998-2001	✓	✓	✓		
1998-2002	✓	✓	✓		
1998-2003	✓	✓	✓		
1998-2004	✓	✓			
1998-2005	✓	✓			
1998-2006	✓	✓			
1998-2008	✓				
2004-2006				✓	✓
2004 & 2008					✓
2005 & 2008					✓

1. Aussi tests cognitifs : poids IPO 1999 à 2001 et IPO 2000-2001.

## *Pondérations disponibles*

- Les pondérations créées couvrent beaucoup de situations d'analyse
- La documentation qui accompagne les fichiers de pondération fournit les explications nécessaires pour l'utilisation

## *Pondérations disponibles*

- Mais,
  - Flexibilité de la collecte implique que les répondants diffèrent selon l'instrument (pas de hiérarchie de réponses dans un volet)
  - Autant de possibilités de poids qu'il y a de combinaisons d'instruments et d'années d'enquêtes
- **Impensable d'avoir une pondération disponible pour toutes les possibilités d'analyses**
  - Approche ISQ : quelques poids créés seulement

## *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 !!**

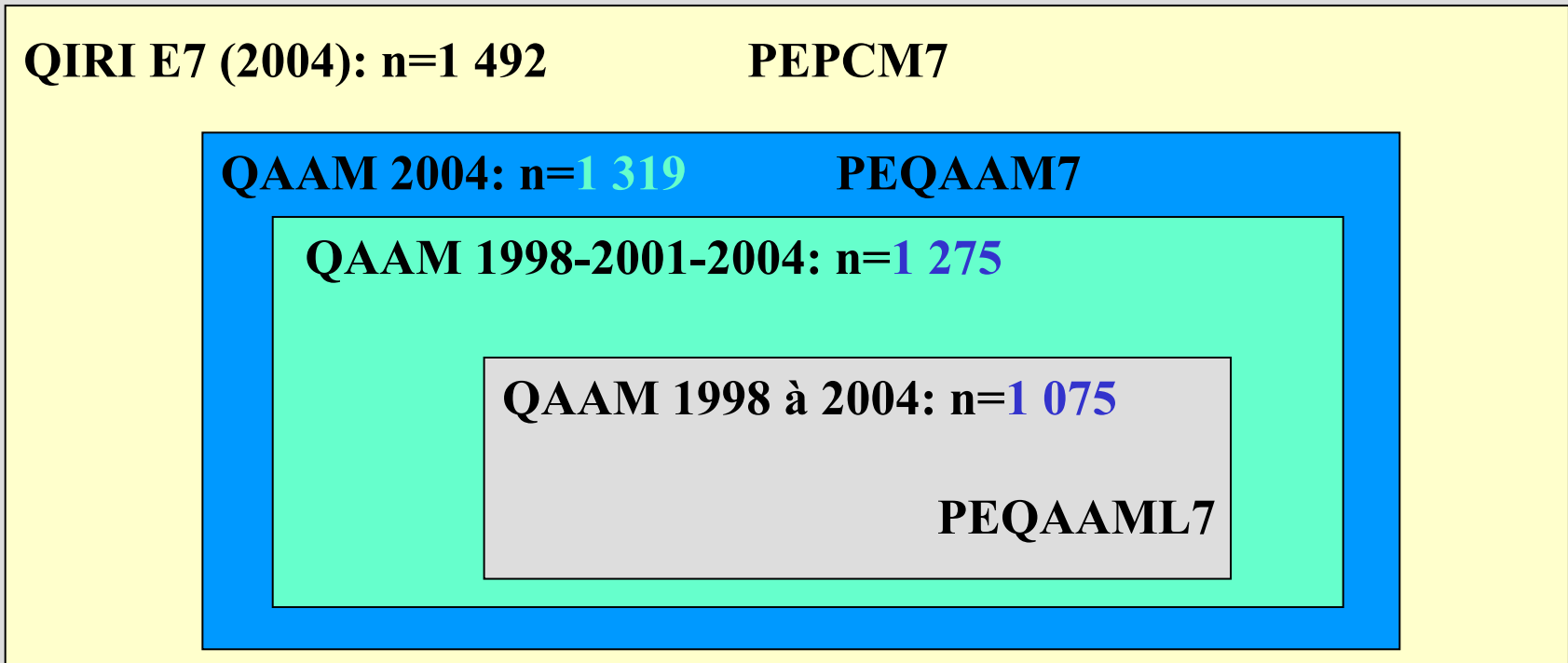
1. Choisir une pondération
2. Calculer la proportion d'unités d'analyse perdues p.c.q. pas de poids
3. Calculer la proportion des non-répondants parmi les unités pour lesquelles un poids existe
4. Évaluer les pondérations alternatives
5. Choisir la meilleure pondération en cherchant à limiter la perte 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 spécifique pour ce besoin
- Possibilités : PEQAAML7 ou PEQAAM7
- Dans les 2 cas, ajustement pour la non-réponse à partir du poids de référence PEPCM7 (QIRI E7)
  - 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?

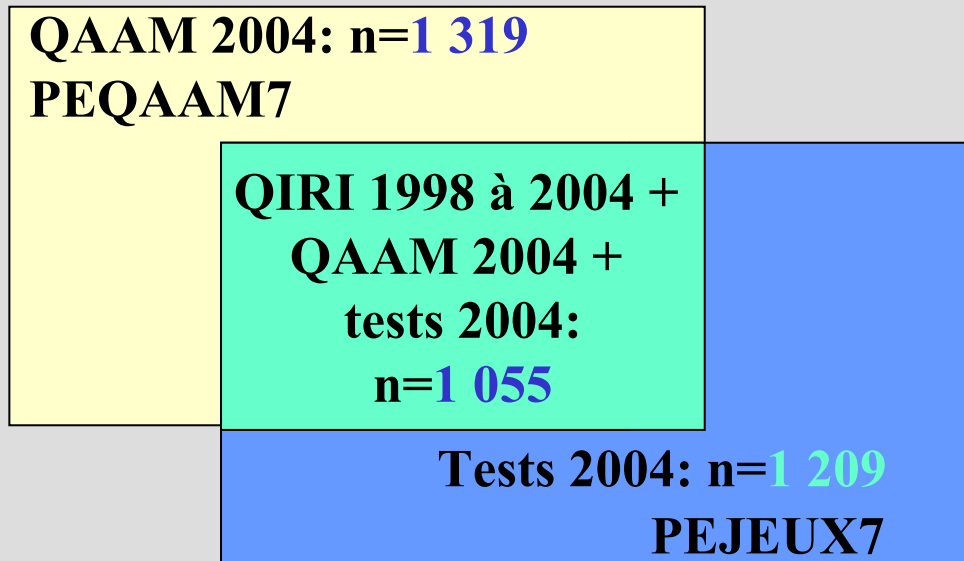


## *Choix d'une pondération*

### *Exemple 2*

- Analyse longitudinale QIRI 1998 à 2004 + QAAM et tests cognitifs 2004
- 1 055 unités d'analyse visées
- PEQAAM7 ou PEJEUX7?
- Dans les 2 cas :
  - Aucune perte d'unités d'analyse

- Unités d'analyse visées (**vert**) : n=1 055
- PEQAAM7 (**jaune**) : ne tient pas compte de 20 % de la non-réponse  
(1 319-1 055)/1 319 → biais?
- PEJEUX7 (**bleu**) : ne tient pas compte de 13 % de la non-réponse  
(1 209-1 055)/1 209 → biais?





## *Comment pondérer?*

- Pondération disponible pour la situation d'analyse envisagée
  - Consulter la documentation du fichier de pondération pour savoir quel poids utiliser
- Pondération non disponible pour la situation d'analyse envisagée
  - Choisir parmi les pondérations disponibles laquelle représente le meilleur compromis
    - Au besoin – consulter l'ISQ
  - Possibilité de pondérations sur demande

## *Non-réponse partielle*

- **Non-réponse à une question donnée parmi les répondants à un instrument**
  - ≠ 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

## *Non-réponse partielle*

- Non-réponse partielle trop élevée
- Traitements pour cette non-réponse possibles (pondération, imputation)
- Mise en garde! Le choix du traitement dépend :
  - % non-réponse partielle
  - Type d'analyse à effectuer
  - Nombre de variables impliquées dans le modèle avec non-réponse partielle

## *Exemple de non-réponse partielle*

### Évaluation de l'enfant par l'enseignant quant aux compétences en 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?

## *Comment tenir compte du plan de sondage?*

- Procédures (logiciels) habituelles
  - Supposent E.A.S.
  - Danger de mauvaises conclusions pour les tests
- Procédures (logiciels) spéciales
  - Tiennent compte du plan stratifié à deux degrés de l'ÉLDEQ
  - Conclusions des tests correctes

## *Plan de sondage complexe* *SAS et SUDAAN*

SAS	SUDAAN	Description
SURVEYMEANS	DESCRIPT / RATIO	Totaux, moyennes, quantiles et ratios
SURVEYFREQ	CROSSTAB	Tableaux de fréquence et tests d'indépendance
SURVEYREG	REGRESS	Régression linéaire
SURVEYLOGISTIC	RLOGIST	Régression logistique
	MULTILOG	Régression logistique multinomiale
	SURVIVAL	Analyse des durées de vie
	LOGLINK	Régression de poisson

## *Plan de sondage complexe SAS et SUDAAN*

- PROC SURVEYREG data=eldeq;  
weight pepcm7;  
strata strate1; cluster upecode;  
MODEL anxise = age;  
run;
- PROC REGRESS data=eldeq filetype=sas design=WR;  
weight pepcm7;  
nest strate1 upecode/ missunit;  
model anxise = age;  
run;

## *Plan de sondage complexe*

### *Autres logiciels*

- SPSS, MPLUS, STATA, LISREL, HLM
- Plusieurs autres logiciels ou procédures de SAS permettent l'utilisation de la pondération
- Souvent, on ne peut pas tenir compte des aspects liés à la précision et aux tests d'hypothèses



## *Comment tenir compte du plan de sondage?*

- Plusieurs logiciels statistiques spécialisés
- Les logiciels « habituels » proposent des procédures spécifiques
- Nécessité de spécifier une variable de pondération
- Nécessité de spécifier les variables décrivant le plan de sondage

## *Conclusion*

- Les analyses non pondérées qui ne tiennent pas compte du plan de sondage risquent de mener à des conclusions fausses
  - Biais dû à la non-réponse
  - Plan de sondage stratifié à deux degrés
  
- Les pondérations disponibles pour faire des analyses justes couvrent plusieurs situations d'analyse
  - Possibilité d'adaptation au besoin
  
- Plusieurs logiciels offrent les outils pour tenir compte de la pondération et du plan de sondage de l'ÉLDEQ



# Merci de votre attention

Site Web de l'ÉLDEQ: [www.jesuisjeserai.stat.gouv.qc.ca](http://www.jesuisjeserai.stat.gouv.qc.ca)



## Références

- Desrosiers, H., V. Dumitru et L. Dubois (2009). « Le surplus de poids chez les enfants de 4 à 7 ans : des cibles pour l'action », dans : *Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2010)*, Québec, Institut de la statistique du Québec, vol. 4, fascicule 3.
- Desrosiers, H., et A. Ducharme (2008). « Poverty, Child Health, and Cognitive Development at age 6: Somme results from the Québec Longitudinal Study of Child Development (QLSCD- Canada) », affiche présentée au 20<sup>e</sup> congrès bisannuel de l'International Society for the Study of Behavioural Development (ISSBD), Würzburg, Allemagne, 15 juillet
- Lemelin, J.-P., et M. Boivin (2007). « Mieux réussir dès la première année : l'importance de la préparation à l'école », dans : *Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2010)*, Québec, Institut de la statistique du Québec, vol. 4, fascicule 2.