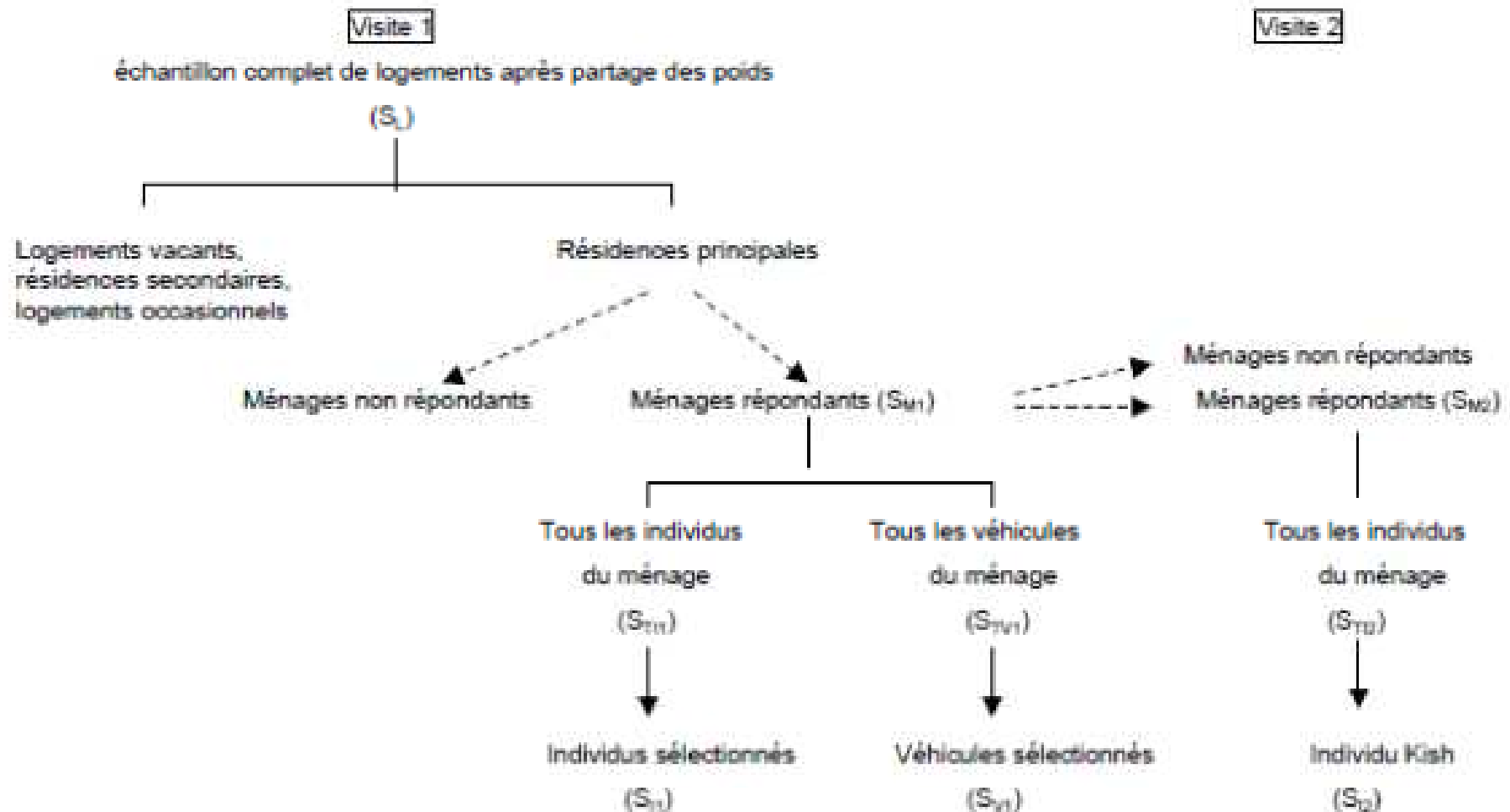


Modélisation sur « petits domaines » pour des estimations régionales en France de métropole

- dans l'enquête nationale sur les transports et déplacements (ENTD 2008)
- dans l'enquête handicap-santé auprès des ménages (HSM 2008)

L'enquête nationale Transports et Déplacements de 2007-2008

- Population de référence : personnes âgées de 6 ans ou plus
- Mesure de la fréquence des déplacements selon :
 - le motif
 - le mode de transport utilisé
 - le jour de la semaine
- Durée des déplacements
- Distances parcourues



Lecture :

Chaque flèche indique une procédure de sélection : \longrightarrow

déterminée par le plan de sondage

\dashrightarrow

déterminée par le comportement de réponse

L'enquête nationale Transports et Déplacements de 2007-2008

- Échantillon de 18600 ménages répondants en métropole
- Ile de France et Pays de Loire : échantillons régionaux exploitables (5400 et 2600 répondants)
- Languedoc-Roussillon, Bretagne, Midi-Pyrénées : un millier de répondants dans chaque région
- Autres régions : estimateurs Horvitz-Thomson directs non publiables

ENTD 2008 : le modèle de comportement

- Y_k =nombre de déplacements du ménage
- X_k =zone de résidence du ménage
(taille d'agglomération * centre ou périphérie)
type de ménage (structure familiale)
nombre de voitures du ménage
CSP du chef de ménage
nombre de personnes du ménage par sexe*âge
nombre de personnes du ménage par CSP

ENTD 2008 : le modèle de comportement

Variable expliquée	R2	
	France de province	Ile de France
Population mobile le jour de référence	0,71	0,76
Nombre de déplacements pour les études	0,57	0,51
Nombre de déplacements par ménage le jour de référence	0,46	0,48
Nombre de déplacements pour le travail ou les études	0,40	0,42
Nombre de déplacements pour le travail habituel	0,39	0,42
Nombre de déplacements en voiture (conducteur)	0,37	0,29
Temps total passé dans les déplacements	0,33	0,39
Distance totale parcourue le jour de référence	0,23	0,32
Nombre de déplacements en transport en commun	0,21	0,21
Nombre de déplacements à pied ou en vélo	0,17	0,24
Nombre de déplacements pour des achats	0,11	0,15
Nombre de déplacements pour les loisirs	0,09	0,13

ENTD 2008 : les méthodes d'estimation testées

- Estimateur par calage sur des totaux régionaux

$$\hat{Y}_{d,cal} \approx \hat{Y}_{d,reg} = \hat{Y}_{d,HT} + \hat{\beta}_d' (\mathbf{X}_d - \hat{\mathbf{X}}_{d,HT})$$

avec : $\hat{\beta}_d = (\mathbf{X}_d' \mathbf{D} \mathbf{X}_d)^{-1} \mathbf{X}_d' \mathbf{D} \mathbf{Y}_d$ $\mathbf{D} = \text{diag} \left(\frac{1}{\pi_i} \right)$

- Estimateur synthétique par régression

$$\hat{Y}_{d,regsyn} = \hat{\mathbf{b}}' \mathbf{X}_d \quad \text{avec :} \quad \hat{\mathbf{b}} = (\mathbf{X}' \mathbf{D} \mathbf{X})^{-1} \mathbf{X}' \mathbf{D} \mathbf{Y}$$

- Estimateur indirect avec modélisation explicite :
modèle avec la région en effet aléatoire

$$Y_{d,i} = \beta' \mathbf{X}_{d,i} + v_d + e_{d,i}$$

$$\hat{Y}_d = \hat{\beta}' \sum_{i \in U_d} \mathbf{x}_i + \hat{v}_d N_d = \hat{\beta}' \mathbf{X}_d + \hat{v}_d N_d$$

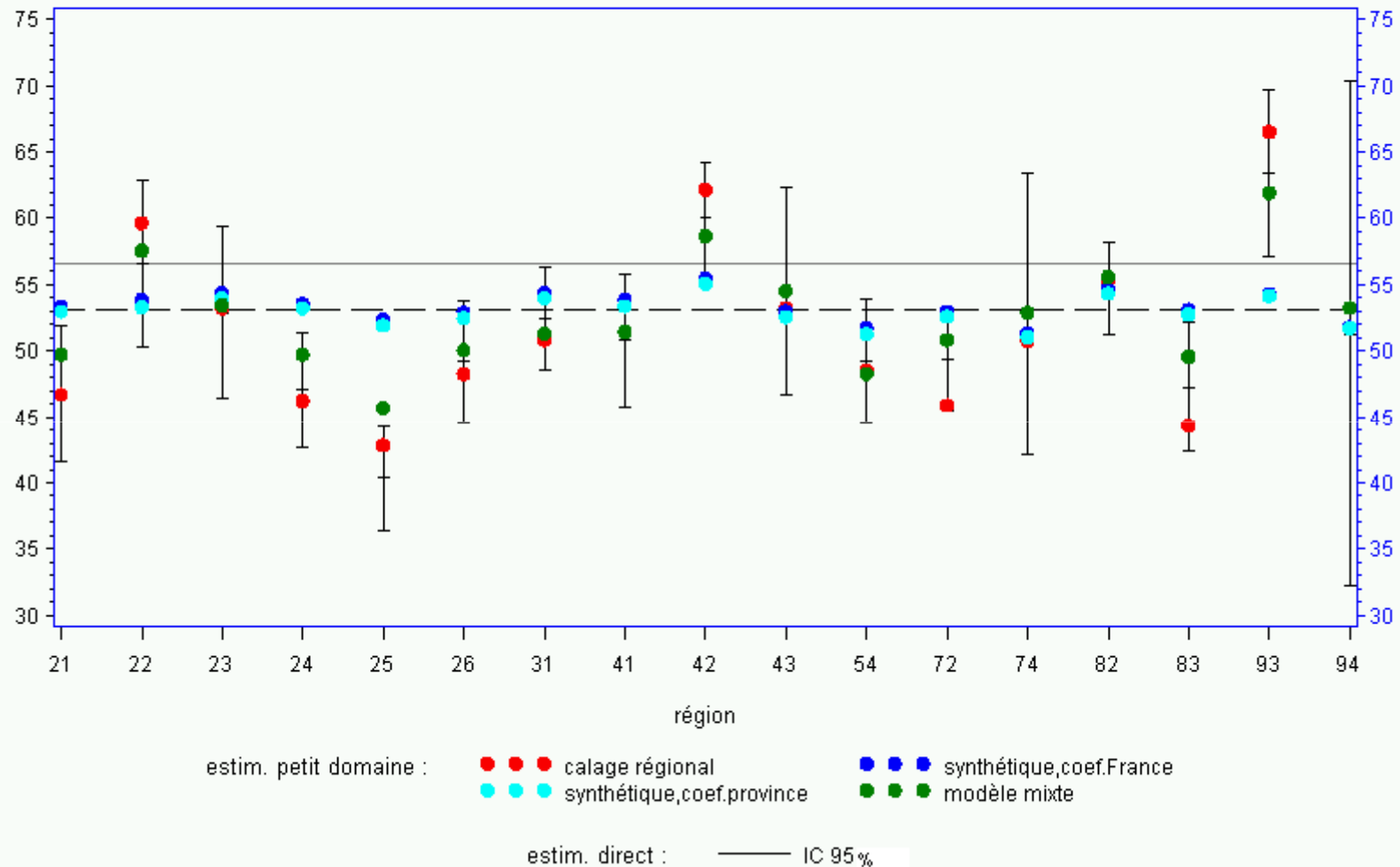
ENTD 2008 : mode de validation des résultats

- Sauf pour l'estimateur direct, pas d'estimateur sans biais de la variance des estimateurs « petits domaines »
- Régions avec extension d'échantillon=zones tests
- Recouvrement de l'estimateur « petit domaine » par l'intervalle de confiance à 95 % de l'estimateur direct

Temps passé dans les déplacements un jour de semaine par habitant

Région	estimateur direct borne inf. de l'IC	estimateur direct (poids nationaux)	estimateur direct borne sup. de l'IC	estimateur synthétique par régression : $Y_r = b * X_r$ b France	estimateur synthétique par régression : $Y_r = a * X_r$ a province	estimateur par modèle mixte effet régional aléatoire
Ile de France	70	72	74	71	.	71
Pays de la Loire	52	55	57	54	53	54
Bretagne	51	55	58	54	53	55
Midi-Pyrénées	51	53	55	54	54	54
Languedoc-Roussillon	51	53	56	53	53	54

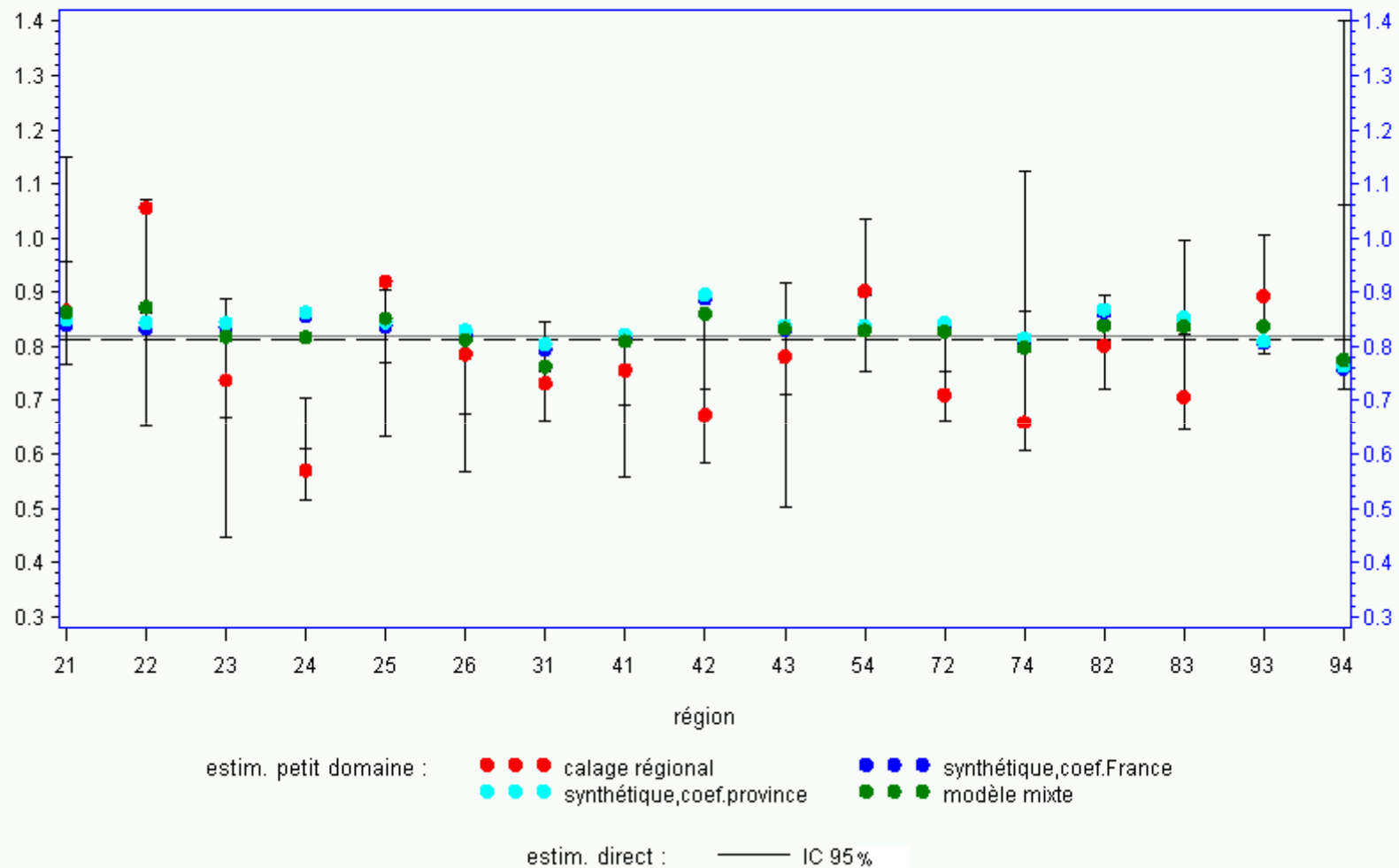
Temps passé dans les déplacements dans la journée par habitant



Déplacements professionnels : fréquence journalière par habitant

Région	estimateur direct borne inf. de l'IC	estimateur direct (poids nationaux)	estimateur direct borne sup. de l'IC	estimateur synthétique par régression : $Y_r = b * X_r$ b France	estimateur synthétique par régression : $Y_r = a * X_r$ a France	estimateur par modèle mixte effet régional aléatoire
Ile de France	0.81	0.84	0.86	0.86	.	0.86
Pays de la Loire	0.95	1.02	1.10	0.89	0.90	0.94
Bretagne	0.77	0.87	0.97	0.86	0.87	0.87
Midi-Pyrénées	0.89	0.95	1.01	0.87	0.87	0.87
Languedoc-Roussillon	0.70	0.76	0.81	0.78	0.78	0.78

Déplacements professionnels : fréquence journalière par habitant



L'enquête Handicap-Santé de 2008

- 2 enquêtes auprès des personnes handicapées :
 - celles vivant à leur domicile
(« ménages ordinaires ») -> HSM
 - celles résidant en institutions -> HSI
- Estimations sur « petits domaines » à partir de la 1^{ère} enquête HSM seule

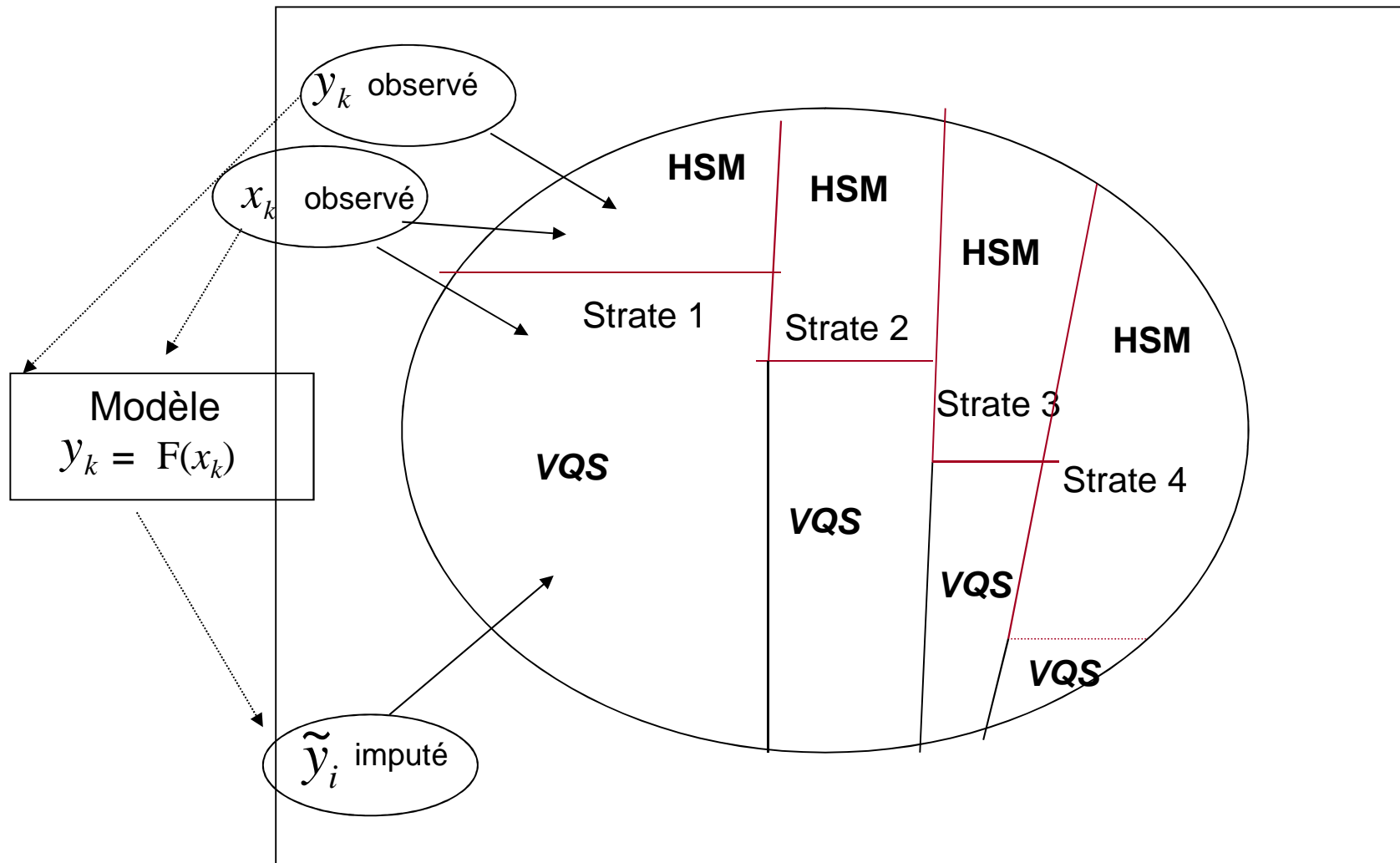
L'enquête HSM 2008

- Enquête en deux phases
- **Première phase : vie quotidienne et santé (VQS)**
 - Objectif : repérer les personnes présumées handicapées
 - Gros échantillon : >218000 personnes répondantes en métropole
 - Voie postale, avec relances par téléphone ou par enquêteur
 - Questionnaire :
 - 1 feuille recto-verso avec 1 colonne par membre du ménage
 - questions simples sur les difficultés ressenties
 - réponses par (oui,non) ou (pas de difficultés, quelques difficultés, beaucoup de difficultés)

L'enquête HSM 2008

- **Deuxième phase HSM :**
 - Échantillon stratifié en 4 groupes construits à partir des réponses au questionnaire VQS, avec taux de sondage croissant de la strate 1 à la strate 4
 - Strate 1 : personnes présumées en bonne santé
 - ...
 - Strate 4 : personnes présumées handicapées
 - Enquête en face à face
 - Questionnaire très détaillé recueillant des données factuelles (la personne fait ou ne fait pas, ...)
 - 26400 personnes répondantes en métropole

La méthode d'estimation sur « petit domaine »



Estimateur de domaine

Deux déclinaisons

- somme des valeurs prédites dans tout VQS :

$$\hat{Y}_{1d} = \frac{1}{\hat{N}_d} \sum_{i \in U_d \cap S_1} w_{1i} \tilde{y}_i$$

- somme dans VQS des prédictions et des valeurs observées dans HSM :

$$\hat{Y}_{2d} = \frac{1}{\hat{N}_d} \left(\sum_{k \in U_d \cap S_2} w_{1k} y_k + \sum_{j \in U_d \cap S_d \cap \bar{S}_2} w_{1j} \tilde{y}_j \right)$$

Estimateur de domaine

- EQM de l'estimateur direct Horvitz-Thomson :

- EQM=variance (estimateur sans biais)

- Variance :

$$Var(\hat{Y}_{HT}) = Var[E(\hat{Y}_{HT} / \mathbf{s}_1)] + E[Var(\hat{Y}_{HT} / \mathbf{s}_1)]$$

- EQM de l'estimateur modélisé :

- EQM=variance + biais²

- Variance = variance de sondage due à l'échantillonnage de première phase (incluant la variance du coefficient de régression, estimé sur échantillon)

=> réduite si la part de la 2^{ème} phase dans la variance totale est importante

Les variables estimées par région et département

- Prévalence du « handicap » (*déficience + altération fonctionnelle connexe*) :
 - par genre :
 - sensoriels (auditifs et visuels)
 - moteurs
 - cognitifs
 - selon l'âge :
 - population de 20 à 59 ans
 - population de 60 ans ou plus
- Prévalence des personnes âgées dépendantes :
 - part de la population de 60 ans et plus appartenant aux GIR 1 à 4

Le modèle

- le modèle estime une probabilité d’être reconnu handicapé ou dépendant dans le questionnaire médical de 2^{ème} phase, en fonction de l’état de santé ressenti déclaré en 1^{ère} phase
- les facteurs « explicatifs » sont recherchés parmi :
 - les réponses au questionnaire VQS de 1^{ère} phase
 - les critères sociodémographiques présents dans la base de sondage ou relevés dans les deux phases
- régression logistique mixte avec le département en effet aléatoire
 - pour l’individu i résidant dans le domaine d :

$$\text{Prob}\{Y_{d,i} = 1\} = p_{d,i} = \frac{1}{1 + \exp[-(\mathbf{b}' \mathbf{x}_i + v_d)]}$$

Le modèle : facteurs fixes

- Les réponses à l'enquête VQS de 1^{ère} phase : quatre groupes de questions
 - difficultés **sensorielles** (voir de près, de loin, entendre)
 - difficultés de **mobilité** (marcher, se baisser, attraper en hauteur, se servir de ses mains)
 - difficultés **intellectuelles** (mémoriser, comprendre, communiquer, prendre des initiatives, être autonome)
 - avoir besoin d'une **aide** technique (prothèse...) ou humaine (aidant régulier) dans la vie quotidienne, **ne pas pouvoir réaliser** une activité, **se sentir handicapé**, avoir une **reconnaissance** officielle d'un handicap

Le modèle : facteurs fixes

- Critères sociodémographiques significatifs :
 - l'âge au-delà de 60 ans
 - le sexe (*handicaps par type des personnes âgées, limitations d'activité*)

Critères de validation des résultats

- 4 départements témoins : Nord, Pas-de-Calais, Rhône, Hauts de Seine, qui ont réalisé une extension d'échantillon à chaque phase
- Comparaison de l'estimateur direct HT à la somme des valeurs prédites sur des niveaux agrégés : France entière, régions
- Distribution des domaines selon l'estimateur HT et l'estimateur modélisé : régression de \hat{Y}_d sur \hat{Y}_{HT}
- Part de la variance de 2ème phase dans la variance totale de l'estimateur direct Horvitz-Thomson

Les résultats

- Les deux méthodes de calcul donnent des résultats presque identiques
- France entière, écart relatif entre prévalence prédite et estimation directe de l'enquête HSM inférieur à 3 % .
L'estimateur prédit est inclus dans l'IC à 95 % de \hat{Y}_{HT} .
- Différence absolue entre la proportion estimée par l'enquête et celle prédite par le modèle inférieure à 1 dans au moins la moitié des régions pour tous les indicateurs calculés, à 2 dans 80 % des régions
- Départements témoins : l'écart relatif entre l'estimation directe et la valeur prédite ne dépasse pas 10 %

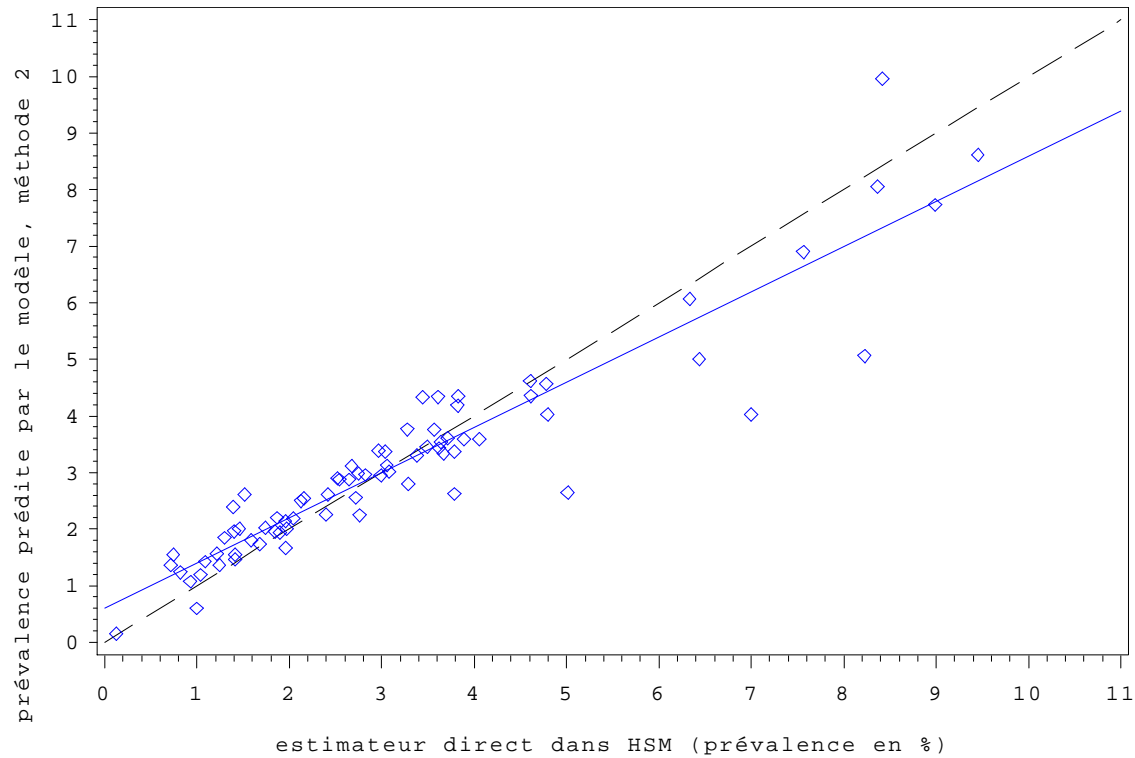
France de métropole

variable	Estimateur direct			Prévalence prédite par le modèle		Écart absolu (2)-(0)	Écart relatif (%) (2)/(0)
	Borne inf. de l'IC	Valeur ponctuelle (0)	Borne sup. de l'IC	méthode 1 (1)	méthode 2 (2)		
<i>Population de 20 à 59 ans</i>							
handicaps cognitifs	4.0	4.5	5.0	4.4	4.5	0.0	0.2
handicaps moteurs	4.1	4.4	4.8	4.3	4.3	-0.1	-3.1
handicaps sensoriels	2.7	3.1	3.5	3.0	3.1	-0.0	-0.2
<i>Population de 60 ans ou plus</i>							
handicaps cognitifs	8.3	9.1	9.9	9.2	9.4	0.3	3.1
handicaps moteurs	23.2	24.4	25.6	24.7	24.8	0.4	1.5
handicaps sensoriels	12.5	13.5	14.4	13.2	13.3	-0.2	-1.2
GIR 1 à 4	6.0	6.5	7.1	6.7	6.7	0.2	3.5

départements témoins : écart entre estimateur direct et estimateur prédit

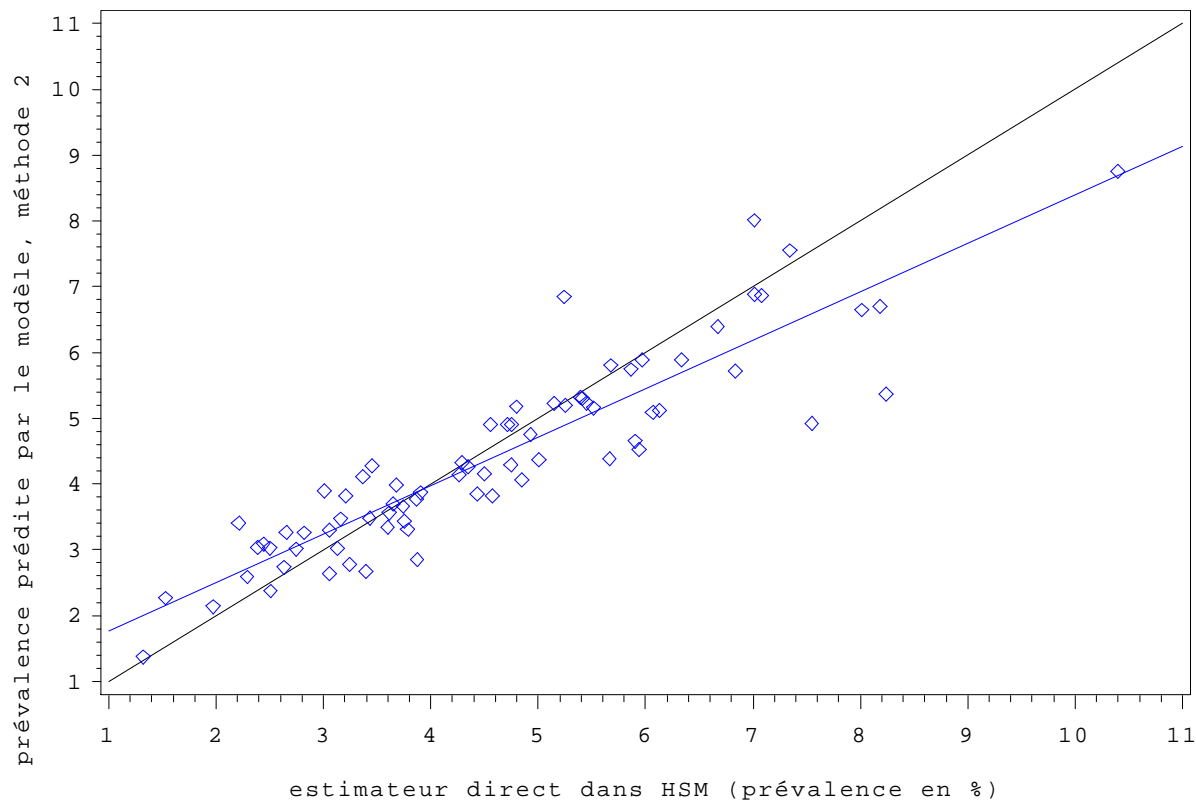
	Nord	Pas de Calais	Rhône	Hauts de Seine
<i>Population de 20 à 59 ans</i>				
altérations sensorielles	-0.2	0.1	0.2	0.2
altérations motrices	0.1	-0.4	0.0	0.3
altérations cognitives	-0.0	0.1	-0.3	-0.1
<i>Population de 60 ans et plus</i>				
altérations sensorielles	0.2	0.6	-0.7	0.6
altérations motrices	0.4	1.2	-0.7	0.8
altérations cognitives	1.1	0.5	-0.1	0.9
GIR 1 à 4	0.8	0.2	-0.1	0.6

Limitations sensorielles dans la population de 20 à 59 ans



Equation de régression : $\text{prévalence prédite} = 0.60 + 0.80 * \text{estimation directe}$

Limitations motrices dans la population de 20 à 59 ans

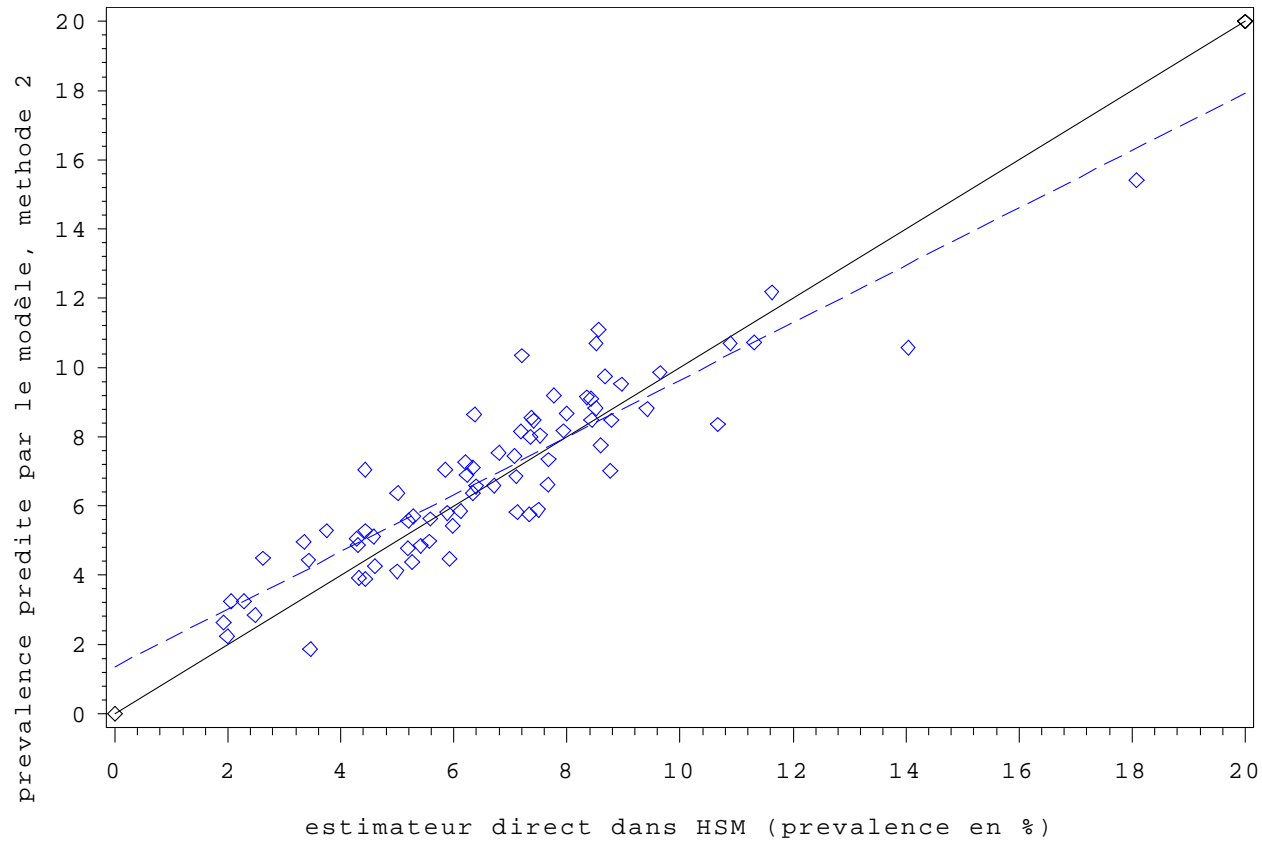


Equation de régression : prévalence prédite = 1.019 + 0.738 *
estimation directe

Sans la Corse et la Loire : prévalence prédite = 0.820 + 0.793 *
estimation directe

GIR 1 à 4

dans la population de 60 ans ou plus



Equation de régression : prévalence prédite = 1.36 + 0.83 * estimation directe

Publication des estimations départementales

Règle de décision :

- Coefficient de variation de l'estimateur direct en phase VQS < 20 % : prévalence départementale prédite publiée sans restriction
- Coefficient de variation de l'estimateur direct en phase VQS compris entre 20 % et 30 % : prévalence départementale prédite publiée avec avertissement
- De 25 (GIR 1 à 4) à 34 (handicaps moteurs des 20-59 ans) départements ont été publiés