

*Décrire les abords et l'état des  
logements pour redresser une  
enquête en face-à-face. L'exemple de  
l'enquête santé ESPS 2014*

Stéphane Legleye, Insee DMCSI

Stéphanie Guillaume, Irdes

Paul Dourgnon, Irdes

# Plan

- Redresser une enquête en deux mots
- Les types de données pour une CNRT
- Description de ESPS
  - Variables, Représentativité
- Procédure pour une CNRT
- Comparaisons des estimations CNRT-redressées et les estimations directement calées
- Discussion des limites et des généralisations

# Redresser une enquête

- Un temps: calage direct
- Deux temps : correction de la non-réponse totale (CNRT) et calage
  - Variables pour une CNRT (Little & Vartivarian, 2005):
    - Présentes pour les non-répondants et les répondants
    - Liées à la participation
    - Liées aux variables cibles de l'enquête ( $\rho > 0,5$ )
    - Sans erreur de mesure ni valeurs manquantes
  - ➔ Variables auxiliaires
  - ➔ Paradoonnées
    - Qualité (Couper, 1998) -> utiles pour l'apurement
    - Level of effort (Biemer, Chen, & Wang, 2013; Olson, 2013)

# ESPS (enquête sur la santé et la protection sociale)

- Panel quadriennal depuis 1988
- Tirage aléatoire simple de ménages dans la base des bénéficiaires de l'assurance maladie
- Tous les membres du ménage participent
- Collecte multimode (face-à-face, papier et téléphone suivant les parties du questionnaire); deux visites en face-à-face (ou deux questionnaires téléphoniques)
- Ici: ESPS 2014, échantillon face-à-face, première visite, uniquement le bénéficiaire tiré
  - Ménages dont le bénéficiaire tiré est âgé ou bénéficiaire de la CMU complémentaire (couverture maladie universelle) ou les ménages dont on n'a pas pu trouver un numéro de téléphone dans l'annuaire

# Variables auxiliaires

- CMU, âge, sexe, numéro de téléphone (raison du face-à-face)
- Données de contexte géolocalisées (IRIS, commune): 57
  - Déciles de PCS ou d'éducation de niveau communal ou IRIS, etc.
- Statut spécial (liées au sujet) et tirées du SNIIR-AM:
  - Nombre de consultations chez un spécialiste
  - Nombre de consultations chez un généraliste
  - Total des dépenses chez les généralistes
  - Total des dépenses chez les spécialistes

# Paradonnées

- Ajout dans les fiches-adresses de:
  - *Logement* : « S'agit-il d'un logement... 1/ Individuel (maison, pavillon) ; 2/ Collectif mais de petite taille (maison avec plusieurs sonnettes) ; 3/ Collectif de grande taille (immeuble, HLM, cité) ».
  - *Allure* : « Quelle est l'allure générale du bâtiment et de ses accès ? 1/ Bien tenu en apparence ; 2/ Quelque peu dégradé mais pas inaccueillant ; 3/ Très dégradé et inhospitalier (très sale, abimé, taggé). »
  - *Hostile* : « Indiquez s'il y a présence d'individus hostiles, menaçant, gênants ou surveillant l'entrée, décourageant l'accès (jeunes, bandes, chiens...). 1/ Oui ; 2/ Non. »
  - Nombre de visites en face-à-face

# Variables cibles

- Etat de de santé (déclaré par les répondants)
  - Se déclarer limité dans les gestes quotidiens
  - Se déclarer en mauvaise santé
  - Se déclarer en affection de longue durée
  - Se déclarer affecté d'une maladie chronique
- 4 variables auxiliaires médico-administratives:
  - Nombre de consultations chez un spécialiste
  - Nombre de consultations chez un généraliste
  - Total des dépenses chez les généralistes
  - Total des dépenses chez les spécialistes

# Calage et variables de calage

- Calage direct simple sur 6 variables :
  - Taille du ménage (1, 2, 3, 4+) ;
  - sexe ;
  - classe d'âge (14 modalités) ;
  - régime d'assurance maladie (4 modalités) ;
  - CMU (2 modalités) ;
  - IdF (2 modalités).



# Procédure

- CNRT fondée sur variables auxiliaires + calage
- CNRT fondée sur parodonnées + calage
- CNRT fondée sur auxiliaires et parodonnées + calage
  
- Comparaison des estimations CNRT/calage aux estimations calées classiques pour :
  - Totaux des 4 Variables médico-administratives
  - Proportions de variables cibles

# Description de l'échantillon

Après élimination des observations sans variables médico-administratives (n=705) et sans paradonnées (n=2435), reste 10272 observations dont 5535 réponses (53,9%)

Paradonnées	%
<b>Logement</b>	
Individuel	65.7
Collectif petite taille	10.3
Collectif de grande taille	24.0
<b>Allure</b>	
Bien tenu en apparence	86.9
Quelque peu dégradé mais pas inaccueillant	11.4
Très dégradé et inhospitalier (très sale, abîmé, taggé)	1.7
<b>Hostile (individus décourageant l'accès)</b>	
Oui	2.7

# Représentativité

Relativement à la base de sondage et aux 6 variables de calage, indicateur R (Bethlehem et al. 2009):

- $R^1=0,98$  (indicateur simple, 6 effets propres)
  - $R^2=0,84$  (21 effets propres et interactions bivariées)
  - ➔ Niveaux très élevés et marges d'améliorations a priori assez faibles
- 
- Déséquilibré uniquement en *Logement* et *IdF*:
    - plus de maisons individuelles et moins de résidents en IdF parmi les répondants

# Choix des variables pour les modèles de CNRT

- On utilise les 4 variables médico-administratives
  - Nombres des consultations spécialistes et généralistes, dépenses ambulatoires et de spécialistes
  - Analyse en composantes principales
  - On retient la composante 1 (55% de variance)
  - On cherche les meilleures combinaisons de variables pour l'expliquer (analyse de variance avec sélection de variables: GLM)
  - ➔ On maximise donc le lien avec la composante
- On fait ensuite la CNRT (régression logistique, méthode des groupes de réponse homogène, sur décile)

# Sélection des variables candidates pour les CNRT

- Lien avec 1<sup>ère</sup> composante ACP ( $F > 3$  ou diff. standardisée  $d > 10$ )
  - Variables auxiliaires
    - Décile de CAPBEP dans la commune et l'IRIS
    - Décile d'agriculteurs exploitants, de cadres, d'employés, d'ouvriers, de retraités et de CDI dans la commune
    - Age et sexe de l'assuré tiré
    - CMU
  - Paradoxonnées
    - Logement
    - Allure
    - Nombre de visites en face-à-face
- Lien avec participation ( $d > 10$ )
  - Décile communal CAPBEP et diplôme du supérieur, agriculteurs, cadres, employés, ouvriers, retraités, inactifs, familles monoparentales
  - Nombre de visites en face-à-face, Logement, IdF, taille unité urbaine

# Modèles retenus

Procédure: Analyse de variance (GLM) sur les variables candidates précédentes + quelques interactions bivariées.  
Sélection Stepwise et validation croisée pour robustesse

CNRT	NB Variables		GLM (sélection) R <sup>2</sup>	Logistique (CNRT) AUC*
Modèle AUX	10	QFMONO age age2 age2*sexe regime age*NUT age*PQ_FAF_r QCS6*NUT QCSUP*QCS2 QCCAPBEP*NUT	0,22	0,57
Modèle PARA	3	LOGEMENT nbvf nbvf*ALLURE	0,01	0,59
Modèle COMP	14	LOGEMENT age age2 age2*sexe regime age*PQ_FAF_r QCSUP*QCS2 NUT*cmu age*PQ_FAF_r*NUT QICAPBEP*QCS5 nbvf2*ALLURE nbvf*nbvf2 HOSTILE*ALLURE nbvf*HOSTILE	0,23	0,60

AUC=aire sous la courbe ROC

# Distribution des poids

Distribution des Poids CNRT/calés proches de celle du calage direct

Variable	N	Min	P1	P99	Max	CV	Sum
Poids calé direct	5535	1222.0	2758.4	26881.9	68038.5	64.0	48759018
pcnrt_aux_cal	5535	1182.2	2724.9	28122.5	61951.6	64.5	48759018
pcnrt_para_cal	5535	891.1	2285.4	32045.5	74058.5	72.6	48759018
pcnrt_comp_cal	5535	869.6	2386.2	28137.8	63816.4	66.4	48759018

# Estimation des 4 variables auxiliaires médico-administratives (ratios)

	Observations	Dépenses ambulatoires	Dépenses spécialistes	Consultations généralistes	Consultations spécialistes
Calage direct	5535	0.894	0.981	0.954	0.992
cnrt_aux_cal	5535	<b>0.895</b>	0.980	0.954	<b>0.995</b>
cnrt_para_cal	5535	0.894	0.980	<b>0.967</b>	<b>0.996</b>
cnrt_comp_cal	5535	<b>0.895</b>	<b>0.982</b>	<b>0.959</b>	<b>0.999</b>
TOTAL VRAI	10272	1.000	1.000	1.000	1.000

En gras: supérieur au calage direct

Améliorations très faibles... et les totaux des dépenses en ambulatoires et des consultations de généralistes restent mal estimés

Le calage direct fait un très bon travail sauf pour les consultations/dépenses de généraliste



# Estimations de l'état de santé déclaré

	Calage direct		CNRT_AUX_CAL			CNRT_PARA_CAL			CNRT_COMP_CAL		
	%	StdErr	%	StdErr	Deff	%	StdErr	Deff	%	StdErr	Deff
Etat de santé dégradé	8.5	0.4	<b>8.5</b>	0.4	1.01	<b>8.7</b>	0.4	1.06	<b>8.6</b>	0.4	1.02
Limitation fonctionnelle	28.1	0.7	<b>28.1</b>	0.7	1.00	<b>28.1</b>	0.7	1.04	<b>28.2</b>	0.7	1.01
Affection de longue durée	22.1	0.6	<b>22.0</b>	0.6	1.00	<b>22.1</b>	0.6	1.04	<b>22.1</b>	0.6	1.01
Maladie chronique	39.2	0.8	<b>39.3</b>	0.8	1.01	<b>39.2</b>	0.8	1.04	<b>39.4</b>	0.8	1.01
Effet totalement négligeable											

# « Robustesse »

- Même analyse mais avec les données déclarées de santé (plutôt que médico-administratives) pour la CNRT
- ➔ Plus orienté vers ce qu'on souhaite mesurer, mais 53% de réponse uniquement
- Résultat identique...

# Synthèse

- Le calage direct d'ESPS apparemment très efficace, mais sous-estime les dépenses et consultations de généralistes
- Pourquoi la CNRT est-elle inefficace ?
  - Faible lien des variables avec les données médico-administratives (ou de santé déclarée)
  - Age, CMU et régime sont très efficaces dans le calage
- Limites:
  - Exclusion initiale d'observations
  - Mauvaise qualité des variables auxiliaires
  - ➔ Répliquer avec un calage plus classique (sans CMU / régime)?
  - ➔ Variables cibles en faible nombre: étudier les comportements à risque (fumer du tabac, boire de l'alcool etc.) ?

# Enseignements?

- Paradoxonnées: améliorer leur qualité!
  - Faible lien des paradoxonnées avec les variables de santé : biais potentiel de non-réponse faible. Rassurant !
    - Corrélation nombre de visites et dépenses ambulatoires=-0,10
  - Donc peu utiles en CNRT
  - En trouver d'autres ?
    - Qualité du contact avec l'enquêté?
    - Historiques des contacts?
- Variables auxiliaires: améliorer leur qualité

Merci !

# Références

- Bethlehem, J., Cobben, F., & schouten, B. (2009). Des indicateurs de la représentativité aux enquêtes. *Techniques d'Enquêtes*(Recueil du symposium 2008 de Statistique Canada), 1-10.
- Little, R. J. A., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey methodology*, 31(2), 161–168.
- Biemer, P. P., Chen, P., & Wang, K. (2013). Using level-of-effort paradata in non-response adjustments with application to field surveys. *Journal of the Royal Statistical Society Series A*, 176(1), 147-168.
- Olson, K. (2013). Paradata for Nonresponse Adjustment. *The Annals of the American Academy of Political and Social Science*, 645(1), 142-170. doi:10.1177/0002716212459475