

L'évaluation anticipée de biais potentiels en cas de difficultés de collecte

Le cas de l'enquête CVS 2013

Emmanuel Gros
Karim Moussallam



Mesurer pour comprendre



L'enquête Cadre de Vie et Sécurité (CVS)

- Enquête annuelle réalisée par l'Insee depuis 2007
- Échantillon de 24 000 ménages
- Objectif : recenser les faits de victimation intervenus l'année précédente.
- Trois types de questionnaire :
 - ✓ questionnaire ménage ;
 - ✓ questionnaire individu ;
 - ✓ interrogation « sous casque » pour les questions les plus sensibles.

CVS 2013 : une collecte perturbée

- CVS 2013 : 1^{ère} enquête collectée dans le cadre des NCEE
 - ↳ collecte perturbée suite à des problèmes ponctuels :
 - ✓ baisse générale des taux de collecte de 10 points de % ;
 - ✓ dégradation non homogène :
 - phénomène plus marqué en Île-de-France ;
 - 8 ZAE totalement orphelines ;
 - 20 ZAE présentent des taux de collecte non nuls mais inférieurs à 25 %.
- ➔ Risque de biais de non-réponse que les procédures de redressements habituelles ne seraient pas en mesure de corriger correctement.

Un évaluation du biais *ex ante* par simulations

- Informations sur le déroulement de la collecte de CVS 2013 disponibles dès le début de l'été, deux mois avant les données d'enquête apurées.
- ↳ Couplé au fait que CVS est une enquête annuelle, ceci a permis de procéder à une évaluation *ex ante* des biais de non-réponse potentiels en procédant par **simulations** :
 - ✓ réplcation, sur les données des enquêtes 2012 et 2011, des conditions de collecte dégradées de CVS 2013 ;
 - ✓ mise en œuvre des procédures de redressements usuelles sur ces échantillons simulés ;
 - ✓ quantification de l'impact sur les estimations de la dégradation des conditions de collecte
- Problème : en 2013, difficile de faire faire directement la part des choses entre la non-réponse et la non-enquête...

Protocole de simulation (1)

➤ Plus précisément, pour l'évaluation *ex ante* du biais de non-réponse 2013 sur CVS 2012, on a procédé comme suit :

① modélisation, par régression logistique, des probabilités de réponse aux enquêtes 2012 et 2013 :

- variables : tranche d'âge de la PR (5), région (3), strate (type d'agglomération sur 5 positions), type de ménage (4), diplôme de la PR (4), nombre de personnes du logement au RP, statut d'occupation du logement (4), nombre de voitures, département de la région de gestion Île-de-France (5) et ZAE de collecte 2013 très perturbée ;
- on récupère en sortie du modèles sur 2012 les probabilités de réponse estimées des logements répondants à CVS 2012 \hat{p}_{2012}^{\log}
- en appliquant le modèle déterminé sur CVS 2013 aux logements répondants à CVS 2012, on estime pour ces logements une probabilité de réponse « à conditions de collecte CVS 2013 » \hat{p}_{2013}^{\log}

Protocole de simulation (2)

② simulation de 10 000 échantillons de répondants CVS 2012 « aux conditions de collecte de CVS 2013 » :

- pour les 8 ZAE orphelines en 2013, tous les logements sont considérés comme non-répondants ;
- pour les autres ZAE, chaque logement est considéré comme répondant avec une probabilité $\hat{p}_{2013}^{\log} / \hat{p}_{2012}^{\log}$ si celle-ci est inférieure à 1 ; sinon on duplique le logement avec une probabilité $\hat{p}_{2013}^{\log} / \hat{p}_{2012}^{\log} - 1$.

③ redressements selon la méthode de CVS 2012 par calage direct ;

④ mesure de l'impact sur les estimations de la dégradation des conditions de collecte, par comparaison avec les estimations issues de l'échantillon CVS 2012 initial.

➤ On a également simulé CVS 2011 « conditions de collecte 2013 », et CVS 2012 « conditions de collecte 2011 ».

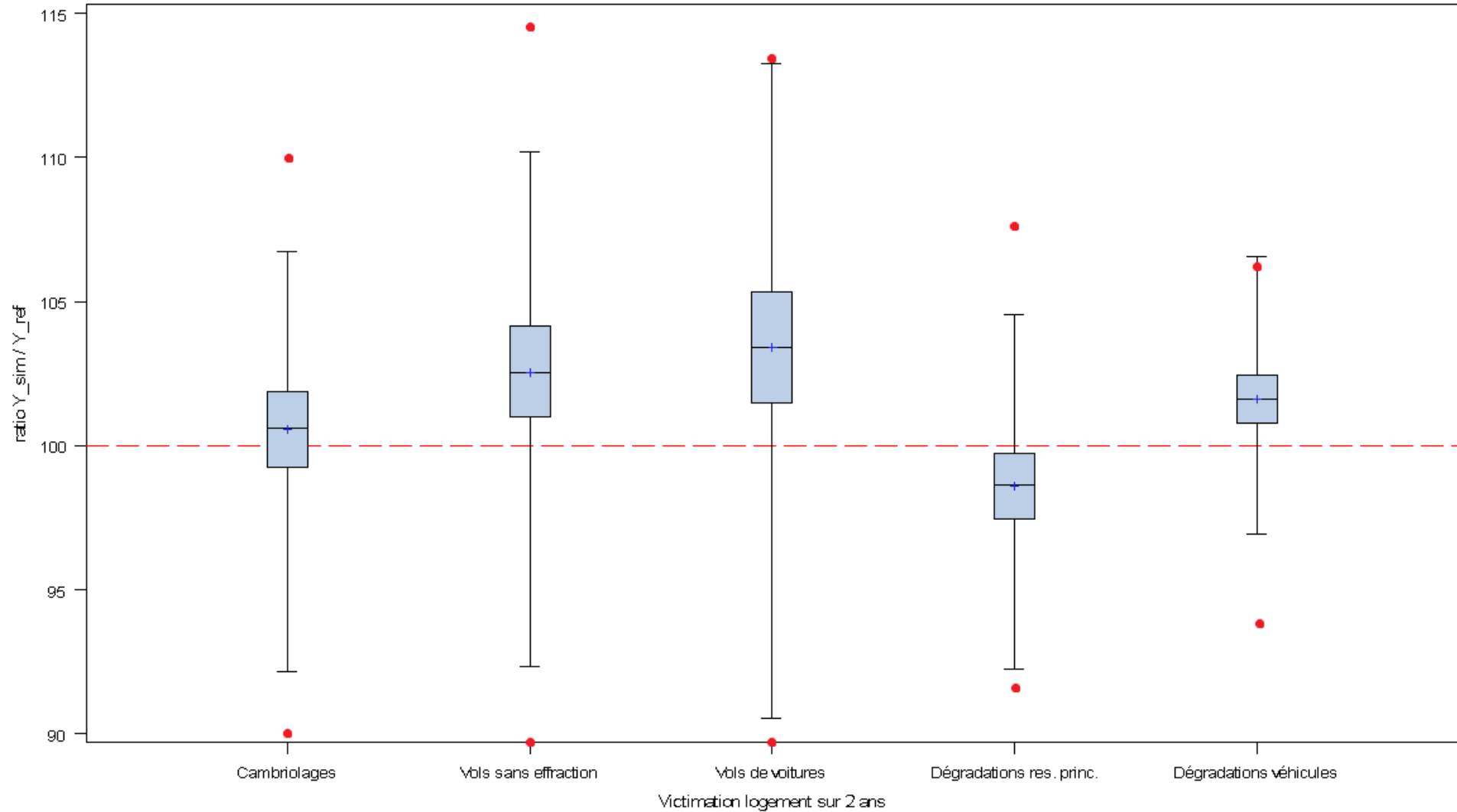
Résultats niveau ménage (1)

- Des biais relatifs anticipés « aux conditions de collecte 2013 » qui restent inférieurs aux coefficients de variation de l'enquête...

	statistique	Nombre de menages	Nombre de personnes du ménage	Cambriolages	Vols sans effraction	Vols résidence secondaire	Vols de voiture	Vols d'objet dans véhicule	Vols de moto	Vols de vélo	Dégradations résidence principale	Dégradations véhicule
CVS 2011 « conditions de collecte 2013 »	CV	0,1	0,1	4,6	6,8	13,3	6,0	3,8	12,2	6,3	4,0	2,8
	Biais relatif	0,0	0,0	-0,4	-0,2	-4,6	4,1	1,5	-0,8	-1,6	0,8	0,9
CVS 2012 : « conditions de collecte 2013 »	CV	0,1	0,1	4,8	6,9	11,4	6,6	3,8	11,6	5,8	4,3	3,1
	Biais relatif	0,0	0,0	0,6	2,5	-0,9	3,4	1,7	0,0	-2,1	-1,4	1,6
« conditions de collecte 2011 »	Biais relatif	0,0	0,0	-0,6	1,1	-1,1	0,4	0,0	2,1	-0,9	0,2	-0,6

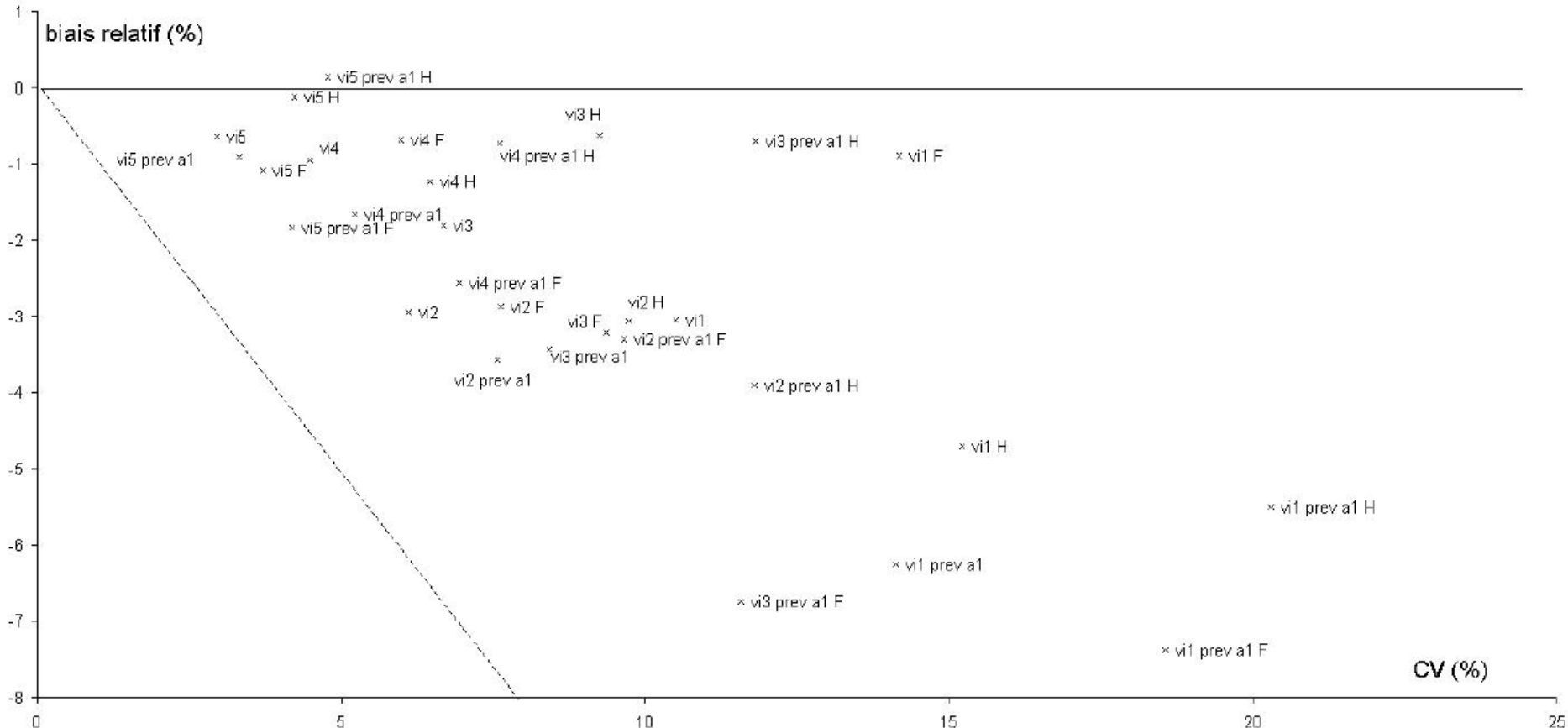
- ... sur 2012 comme sur 2011...
- ... et du même ordre de grandeur que ceux « aux conditions de collecte 2011 ».

Résultats niveau ménage (2)



Résultats niveau individu (1)

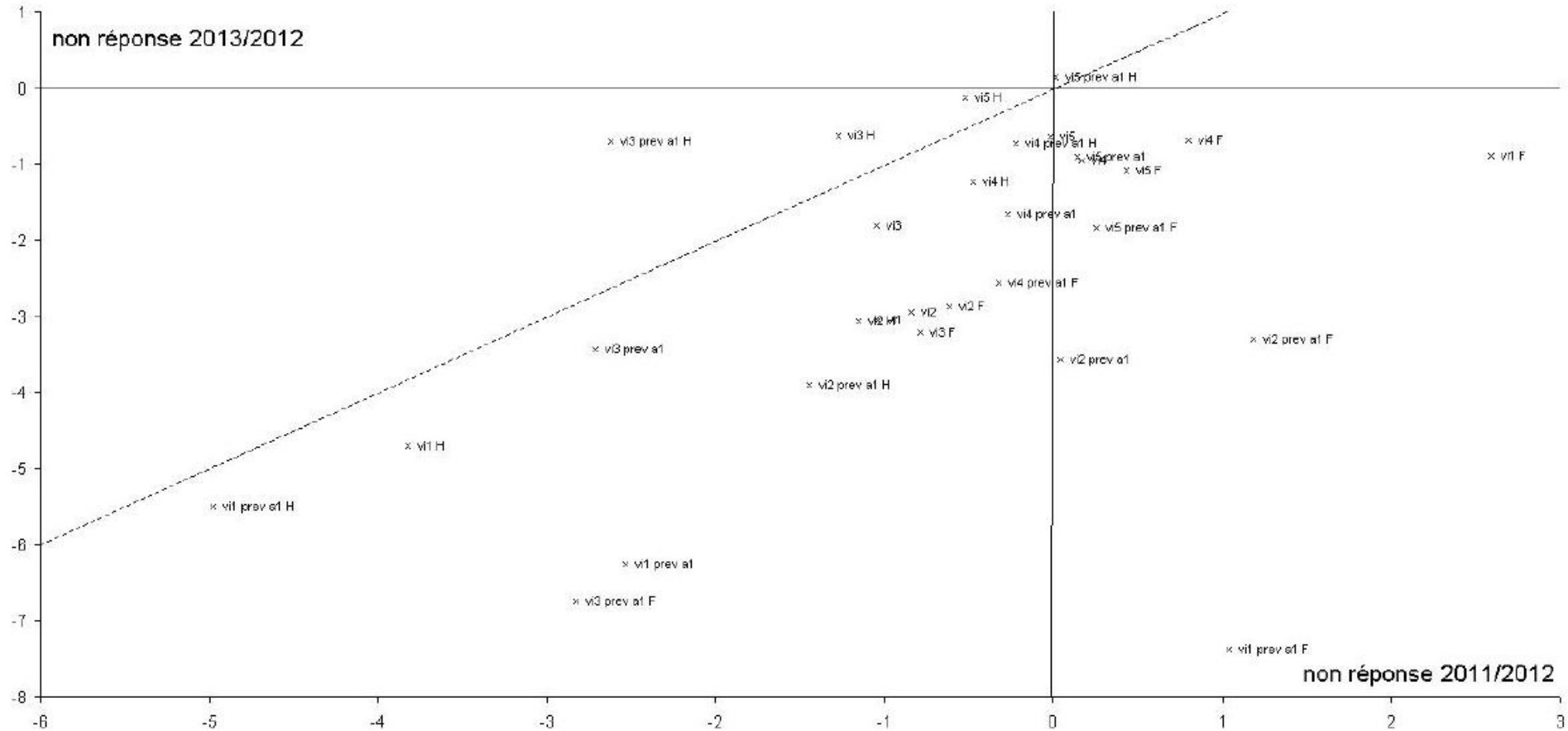
- Des biais relatifs anticipés « 2012 aux conditions de collecte 2013 » toujours inférieurs aux CV...



- ... mais majoritairement négatifs (> 70%).

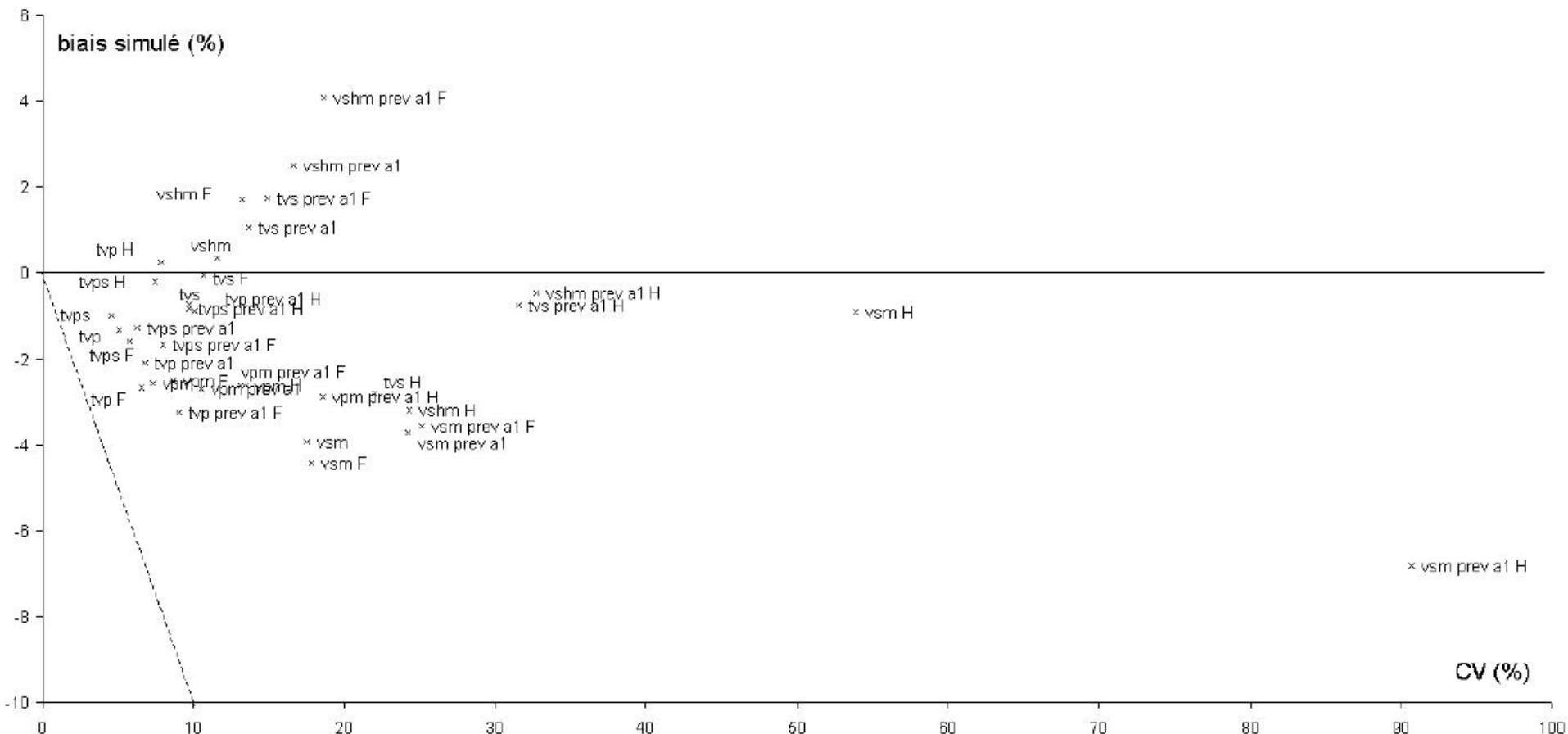
Résultats niveau individu (2)

- Toutefois, on observe aussi ce phénomène pour « CVS 2012 aux conditions de collecte 2011 »



Résultats niveau « casque » (1)

- Des biais relatifs anticipés « CVS 2012 aux conditions de collecte 2013 » encore et toujours inférieurs aux CV...



Résultats niveau « casque » (3)

- En outre, ces variables – violences sexuelles, violences physiques au sein du ménage... – sont particulièrement sensibles...
 - ... et ne concernent que des effectifs très réduits → pas exploitées annuellement, mais en « cumul » sur plusieurs années.
- ➔ Toutes ces raisons ont conduit à ne pas diffuser fin 2013 les variables sous casque, qui font l'objet d'investigations méthodologiques supplémentaires.

Conclusions

- Dans le cadre d'enquêtes répétées dans le temps, des simulations sur données passées permettent d'évaluer *ex ante* de possibles problèmes de biais de non-réponse liés à un problème de collecte.
- Pour CVS 2013 :
 - ✓ pour les principales variables des questionnaires ménages et individus, les simulations n'ont pas révélé l'existence de biais significatifs ;
 - ✓ pour les questions les plus sensibles passées « sous casque », l'ampleur relativement plus importante des biais relatifs « aux conditions de collecte CVS 2013 » par rapport à ceux « aux conditions de collecte CVS 2011 », couplée à la faiblesse des effectifs, ont conduit à ne pas diffuser les résultats → investigations méthodologiques supplémentaires en cours.

Merci de votre attention !

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :

www.insee.fr / Contacter l'Insee

09 72 72 4000

(coût d'un appel local)

du lundi au vendredi de 9h00 à 17h00