

**ENQUÊTE 2012
SUR L'UTILISATION
DES DEUX-ROUES MOTORISÉS**

2RM 2012

RAPPORT QUALITÉ

CGDD / SOeS

Mars 2013

RAPPORT QUALITÉ - Enquête 2RM 2012

Fiche descriptive de l'enquête

« Carte d'identité »	
nom : Enquête 2012 sur l'utilisation des deux-roues motorisés	
n° visa : 2012 X055 TR	
année : 2012	périodicité : quinquennale <i>a priori</i>
service enquêteur responsable SOeS/SDST (Sous-direction Statistiques des transports)	
mode de collecte collecte postale et électronique	

Caractéristiques

Définition du champ :

L'enquête de 2012 sur l'utilisation des deux-roues motorisés (2RM2012) a pour but de déterminer un niveau de parc circulant de deux-roues en France, de mesurer le kilométrage effectué, et de caractériser les usages.

Les véhicules faisant partie du champ de l'enquête sont de type cyclomoteurs ou motocyclettes (tous types de cylindrées), y compris les tricycles pendulaires de type « MP3 ». Les tricycles non pendulaires (de type « Trike »), les quadricycles (de type « Quad »), les side-cars ainsi que tous les véhicules de transport de marchandises ne font pas partie du champ de l'enquête. L'unité statistique est le véhicule enquêté (le deux-roues motorisé). L'enquête porte sur les deux-roues immatriculés en France métropolitaine et dans les 5 DOM (Martinique, Guadeloupe, Guyane, Réunion, Mayotte).

L'enquête est effectuée par sondage dans Sidiv (Système d'information décisionnelle sur les immatriculations de véhicule), qui rassemble l'ensemble des cartes grises (plus de 7 millions de deux-roues en 2013). Une grande partie de ces véhicules, dont les dates de mises en circulation débutent *grosso modo* en 1925, n'est évidemment plus en circulation : véhicules détruits, périmés, hors d'usage, oubliés, exportés, ... Contrairement à certaines autres enquêtes véhicules, comme l'enquête VUL 2011, où un parc de référence est constitué par appariement entre Sidiv et la base des contrôles techniques, un parc de deux-roues ne peut être constitué de cette manière (absence de contrôles techniques).

Par ailleurs, les véhicules appartenant aux entreprises ne sont pas tous conservés dans la base de sondage : sont retirés les véhicules appartenant aux garages, concessionnaires, assurances (rachat de véhicules endommagés ou volés – le professionnel qui achète le véhicule procède alors à sa revente ou à sa destruction), identifiés après appariement avec la base SIRENE. Ne sont conservés que les véhicules susceptibles de participer à la constitution d'un parc « circulant ». Moins de 3 % des deux-roues présents dans la base de sondage, comme dans l'échantillon puis la base finale des répondants, appartiennent à des entreprises. Ces 100 000 véhicules ne disposent pas tous d'un conducteur principal : les résultats montrent qu'un sur trois est conduit par plusieurs conducteurs, sans conducteur principal (coursiers, ...).

Collecte par internet :

L'enquête a été réalisée par voie postale et électronique, en mars 2012.

Unité statistique : véhicule

Un questionnaire est envoyé au titulaire de la carte grise, qui renseigne le début du questionnaire. La suite du questionnaire est alors renseignée par le conducteur principal du véhicule, ou par toute personne habilitée en cas d'absence de ce dernier.

Taille du parc en service : inconnue ex-ante **Taille de l'échantillon :** 30 500
(30 000 2RM de moins de 30 ans, et 500 2RM de plus de 30 ans)

Temps de réponse moyen : 14,5 minutes ; **temps de réponse médian :** 10,0 minutes

Caractère obligatoire : oui (mais pas d'envoi de mises en demeure de répondre)

Les dates-clés

	Prévu	Réalisé
Réunion de la formation CNIS	1 ^{er} semestre 2011	12/05/2011 (avis d'opportunité)
Réunion du Comité du label (et obtention de l'avis de conformité)	Décembre 2011	13 décembre 2011 (consultation écrite)
Déclaration à la CNIL (pour les entrepreneurs individuels et les particuliers selon la norme n°18 pour les EI et la norme n°19 pour les particuliers)	Juin 2011	28 juin 2011 (Récépissé de déclaration normale)
Lancement de la collecte	7 mars 2012	14 mars 2012
Fin de la saisie (via marché de saisie)	Septembre 2012	05 septembre 2012
Base de données finale	Décembre 2012	13 novembre 2012
Retour d'information aux entreprises et particuliers	Non	Non
Mise à disposition du fichier	Mars/Avril 2013	Mars 2013
Diffusion des premiers résultats	Mars 2013	Mars 2013
Diffusion de tableaux détaillés	Mars 2013	Mars 2013

La préparation de l'enquête

existe-t-il un comité d'utilisateurs des résultats de l'enquête ?	Oui	<input checked="" type="checkbox"/> avec utilisateurs externes
nombre de particuliers ayant fait l'objet d'un test de questionnaire	30	
l'échantillon a-t-il fait l'objet d'une démarche de coordination ?	non	

La FFMC (Fédération française des Motards en colère), membre du comité de concertation, a également fait des communiqués pour parler de l'enquête et l'annoncer.

L'échantillon enquêté

Type de sondage

Sondage aléatoire simple stratifié pour les cyclomoteurs et les motocyclettes de moins de 5 ans
Sondage aléatoire à probabilités inégales pour les motocyclettes de plus de 5 ans (cf. **Annexe 1**).

Plan de sondage

Le plan de sondage est détaillé en **Annexe 1**.

Le tirage stratifié repose sur les croisements de trois variables : le type de cylindrée, le zonage urbain, la classe d'âge du véhicule. En outre, pour les véhicules de plus de 5 ans, l'existence du véhicule confirmée ou non dans la base de sondage (Sidiv) détermine une stratification supplémentaire. Au total, 52 strates ont été constituées (Annexe 1).

Le tirage est effectué dans Sidiv (Système d'information décisionnelle sur les immatriculations de véhicules), qui rassemble l'ensemble des cartes grises des véhicules.

Taux de sondage

Les taux de sondage résultent :

- d'un calcul d'optimisation au sens de Neyman dans un premier temps ;
- ensuite d'une « sur-représentation » des motocyclettes dont l'existence est confirmée après le 1^{er} janvier 2008 dans Sidiv, par rapport à celles qui ne le sont pas (la probabilité de tirage dans chaque sous-strate de motocyclettes dont l'existence est confirmée après le 1^{er} janvier 2008 est plus élevée).

Pour les véhicules de moins de 30 ans, la taille de l'échantillon de l'enquête est de 30 000 véhicules, et la population mère est de 6 026 043 véhicules, soit un taux de sondage global de 0,50 % (**Annexe 1**).

Cet échantillon est complété par un sondage stratifié simple de véhicules âgés de 30 ans ou plus, comportant 500 unités, tirées parmi les 125 023 véhicules de 30 ans ou plus dont l'existence est confirmée dans Sidiv après 2008, soit un taux de sondage moyen de 0,40 %.

Distribution des taux de sondage par strate de sondage

Les taux de sondage sont égaux à l'inverse des poids de sondage, dans les 52 strates suivantes (cf. **Annexe 1**) :

Poids de sondage dans chacune des 52 strates (arrondis à l'unité)

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NON CONF.	CONFIRMÉ	NON CONF.	CONFIRMÉ		NON CONF.	CONFIRMÉ	NON CONF.	CONFIRMÉ		NON CONF.	CONFIRMÉ	NON CONF.	CONFIRMÉ	
Cyclo	363	934				135	681				488	415				
Moto 125	138	235	107	784	261	104	191	91	577	206	575	671	335	196	73	
Moto 125-750	75	186	98	671	231	173	215	143	511	196	169	476	317	935	374	
Moto >750	95	188	90	367	131	112	252	148	500	200	144	279	174	175	73	

Les taux de sondage varient de 0,11 % à 1,36 % selon les strates.

Indicateurs liés à la production

y a-t-il des relances ?	Oui
si oui, de quel type et combien ?	Une seule ; relance par courrier, env. 15 000 unités
Nombre d'unités de l'échantillon initial	30 000 (véhicules de moins de 30 ans)
Nombre de retours	14 887
Nombre de NPAI	4 146
Nombre de « hors champ » (y compris unités disparues) dans les retours	néant
Du fait d'un travail d'apurement de la base de sondage, la base n'est constituée que de deux-roues motorisés appartenant au champ de l'enquête. Les véhicules non possédés, détruits, revendus, mis à la casse, ... par les titulaires des cartes grises échantillonnées ne constituent pas des véhicules « hors-champ » : ces derniers sont identifiés grâce aux réponses fournies en début de questionnaire, et leur nombre permet de redresser les unités composant le parc (cf. Annexe 2).	
Nombre de questionnaires exploitables	14 887

Taux de réponse à l'enquête : 49,6 %

Taux de NPAI (retour à l'envoyeur) : 13,8 %

Taux de non-répondants : 36,6 %

Redressement de la non-réponse totale

Une méthode dite de « Modèle de Réponse » a permis de redresser les poids de sondage initiaux. Une comparaison avec un modèle de groupes homogènes de répondants a également été menée en parallèle, à fin de comparaisons. Ces méthodes et les résultats qui en découlent sont décrits en **Annexe 2**.

Redressement de la non-réponse partielle

Le redressement de la non-réponse partielle a porté sur une vingtaine de variables. Selon la nature de la variable, le redressement de la non réponse partielle a été réalisé suivant plusieurs logiques, décrites en **Annexe 3** :

- méthode déductive
- méthode « hotdeck »
- méthode d'imputation par la moyenne

Pourcentage de non-réponses partielles, pour les variables qui ont été redressées

Non-réponse partielle pour les variables qui ont été redressées

Variable	En %
Type de véhicule	0,9
Equipement d'un ABS	1,8
Motorisation du véhicule	1,6
Cycle du moteur essence	5,1
Assurance (au tiers, tous risques) du véhicule	1,5

Assurance contre vol et incendie	3,2
Assurance individuelle	6,4
Kilométrage annuel	1,1
Part kilométrage réalisée par le conducteur principal	4,9
Part du kilométrage réalisée en ville route et autoroute	3,0
Lieu de stationnement du véhicule au domicile	0,7
Période d'utilisation du véhicule	0,4
Fréquence d'utilisation du véhicule	0,6
Utilisation du véhicule la nuit	1,1
Utilisation du véhicule avec un passager	1,3
Fréquence d'utilisation du véhicule pour se rendre au travail	2,1
Utilisation des transports en commun pour se rendre au travail	2,5
Lieu de stationnement du véhicule sur le lieu de travail	4,1
Sexe du conducteur principal	1,0

(cf. **Annexe 3**).

Indicateurs de précision

En appliquant la formule générale du calcul de coefficient de variation dans le cadre d'un sondage stratifié simple, le coefficient de variation de l'estimation du kilométrage moyen est de 0,0202 (2,02 %).

La diffusion

l'enquête a-t-elle fait l'objet d'une procédure d'archivage ?	Elle n'est pas encore réalisée
existe-t-il une documentation pour la diffusion ? Oui	
aucune documentation <input type="checkbox"/>	une documentation simple <input checked="" type="checkbox"/>
un dictionnaire de données complet <input checked="" type="checkbox"/>	une documentation à usage externe <input checked="" type="checkbox"/>
nombre de publications, par type :	
➤ type « Infos Rapides »	1 publication (« Chiffres et Statistiques »)
➤ type « 4 Pages »	1 publication (« Le Point sur »)
➤ type CD-Rom, « Insee-Résultats », web	1 mise en ligne de tableaux de résultats détaillés 1 mise en ligne de fichier détail avec documentation
➤ articles, études	Néant à la date de réalisation de ce bilan

Moyens relatifs à l'opération

Taille de l'équipe (en équivalent temps plein) :
- équipe « statistique » :
0,2 ETP de décembre 2010 à février 2011 ;
0,7 ETP de mars 2011 à février 2012 ;
1,0 ETP de mars 2012 à août 2012 ;
1,7 ETP de septembre 2012 à janvier 2013 ;
0,8 ETP de février 2013 à mars 2013.
- gestionnaire :
0 (le responsable de l'enquête a assuré la gestion de la collecte...)
Moyens informatiques (nombre de jours de développement, maintenance dans le cas d'une opération régulière) :
10 jours pour la préparation et la validation du questionnaire électronique, en lien avec l'équipe informatique
Budget d'environ 15 K€ dédié à la sous-traitance par la réalisation du questionnaire électronique.

Annexe 1

Plan de sondage

Dans cette enquête, l'*unité statistique* échantillonnée est le deux-roues motorisé (cyclomoteur ou motocyclette). Le questionnaire est envoyé au titulaire de la carte grise du véhicule, dont les coordonnées postales figurent sur la carte grise. Le titulaire répond alors à la première question du questionnaire (« question préalable »). Ensuite, le questionnaire permet d'identifier (ou non) le conducteur principal du véhicule ; c'est alors le conducteur principal qui répond à la suite du questionnaire. La population *interrogée* est donc constituée de l'ensemble des titulaires des cartes grises échantillonnées, et la population *concernée* par l'enquête est celle des conducteurs de ces véhicules.

1. Description du sondage

Le sondage s'effectue dans la base des cartes grises (parc FCA/SIDIV). Une grande partie de ce parc, dont les dates de mises en circulation débutent *grosso modo* en 1925, n'est évidemment plus en circulation : véhicules détruits, périmés, hors d'usage, oubliés, exportés, ... Ce parc surestime donc nettement le parc circulant évalué à partir de l'ENTD 2008, l'enquête nationale *Transports et déplacements*.

Le parc circulant estimé à partir de l'ENTD 2008 s'élève quant à lui à 2,99 millions d'unités (0,97 million de cyclomoteurs et 2,02 millions de motos). Ce parc, qui suscite parfois quelques interrogations de la part des partenaires et institutions, ne sert en aucun cas de « calage » pour l'enquête – dont l'un des objectifs vise précisément à déterminer un parc circulant. Néanmoins, l'ENTD va servir de source auxiliaire pour la détermination du plan de sondage, car elle fournit des macro-informations robustes.

1.1. Apurement de la base de sondage (base des cartes grises)

Parmi les deux-roues présents dans la base au sens de la nomenclature des genres et carrosseries des véhicules affectés au transport, sont retirés les tricycles non pendulaires (de type « Trike »), les quadricycles (de type « Quad »), les side-cars ainsi que tous les véhicules de transport de marchandises. Les tricycles pendulaires de type « MP3 » sont en revanche conservés dans la base. Il n'y a pas, après ce premier apurement, de véhicules qui seraient « hors-champ » (voitures, camions, ...), en raison de contrôles multiples sur les types de carrosserie, les genres, mais aussi les marques et les modèles dans certains cas.

Les véhicules appartenant aux entreprises, présents également dans la base des cartes grises, ne sont pas tous conservés : sont retirés de la base les véhicules appartenant aux garages, concessionnaires, assurances (rachat de véhicules endommagés ou volés – le professionnel qui achète le véhicule procède alors à sa revente ou à sa destruction). La base a été appariée avec la base SIRENE : les véhicules associés à un numéro SIREN invalide ou inexistant sont retirés, ainsi que les véhicules appartenant aux secteurs s'activant suivants :

- 3831Z Démantèlement d'épaves
- 4677Z Commerce de gros (commerce interentreprises) de déchets et débris
- 4511Z Commerce de voitures et de véhicules automobiles légers
- 4520A Entretien et réparation de véhicules automobiles légers
- 4532Z Commerce de détail d'équipements automobiles
- 4540Z Commerce et réparation de motocycles
- 6512Z Autres assurances

Les entreprises de ces secteurs sont en effet détentrices d'un nombre important de véhicules, non pas pour leur usage, mais pour leur destruction ou pour leur commercialisation. Ces véhicules ne font pas partie du parc circulant. Les deux premiers secteurs (3831Z et 4677Z) correspondent à des véhicules détruits ou hors d'usage ; les quatre suivants à des véhicules stockés avant mise en vente (non roulants ; essentiellement des garages et des concessionnaires). Le dernier (6512Z) est constitué d'assureurs qui reprennent des épaves après accident avant de les revendre à la casse, ou des véhicules volés de leurs assurés

(« déclaration d'achat du véhicule ») qu'ils remettront ensuite sur le marché, par l'intermédiaire de concessionnaires (« déclaration de cession du véhicule »).

In fine, 2,7 % des deux-roues présents dans la base de sondage appartiennent à des entreprises.

Enfin, les véhicules de plus de 30 ans au 01/01/2012 sont retirés. Seuls les véhicules de moins de 30 ans seront interrogés : cette limite d'âge résulte des comparaisons menées avec l'ENTD 2008 : le nombre de véhicules de plus de 30 ans en circulation est très faible. Toutefois, le budget le permettant, un tirage spécifique parmi les seuls véhicules de plus de 30 ans a été mené, selon les mêmes strates décrites plus loin (500 unités).

Au final, la base de sondage des cartes grises après apurements comporte 6,0 millions de véhicules de moins de 30 ans au 1^{er} janvier 2012 (tableau 1).

1.2. Détermination des strates

À partir de l'ENTD 2008, des analyses de variance ont été menées afin de déterminer les meilleures variables de stratification possibles, parmi celles présentes à la fois dans le parc des cartes grises – base de sondage – et dans l'ENTD, puis les meilleurs seuils optimaux possibles des modalités de chacune de ces variables, en modélisant le logarithme des kilomètres parcourus (distribution rendue normale par cette manipulation).

Suite à cette analyse, trois variables ont été retenues pour la stratification :

- l'âge du véhicule (3 modalités : moins de 5 ans, de 5 à 10 ans, de 11 à 29 ans) ;
- le type de cylindrée (4 modalités : cyclomoteurs, motos 125 cm³, motos 125 à 750 cm³, motos > 750 cm³) ;
- le type de zonage urbain (3 modalités, détaillées ci-après).

Les modalités du zonage urbain retenues ont été élaborées à partir de la nomenclature récente établie par l'Insee – rendue publique au 4^{ème} trimestre 2011 –, mais éclatée puis réassemblée en d'autres sous-groupes, afin d'être représentative en termes d'analyse de variance. Les 3 modalités sont :

- « Aire urbaine de Paris »
Cette modalité regroupe Paris intra-muros (ville-centre), la banlieue du pôle urbain de Paris et la couronne de l'aire urbaine de Paris.
- « Aire urbaine des grands pôles »
Cette modalité regroupe les banlieues, villes isolées et villes-centre ainsi que la couronne des grands pôles (plus de 10 000 emplois), hors aire urbaine de Paris.
- « Petite et moyenne aire urbaine + Multipolarisé + Rural isolé »
Cette modalité rassemble 7 groupes de la nomenclature Insee d'origine : multipolarisé des grands pôles urbains, moyens pôles urbains (5 000 à 10 000 emplois), couronne des moyens pôles urbains, petits pôles urbains (moins de 5 000 emplois), couronne des petits pôles urbains, autre multipolarisé, communes isolées hors influence des pôles urbains.

Les deux premières modalités sont issues de l'éclatement de 2 groupes de la nomenclature Insee sur les zonages, à savoir les grands pôles (plus de 10 000 emplois) et la couronne des grands pôles.

Au final, il y a donc 3 x 4 x 3 = 36 strates. Les effectifs dans la base de sondage des cartes grises après apurements se répartissent alors de la manière suivante au 1^{er} janvier 2012 :

Tableau 1

Effectifs strates BASE de sondage CARTES GRISES											
N_h^{CG} (Carte grise)	Aire urbaine de Paris			Aire urbaine des grands pôles			Petite et moyenne aire urbaine + Multipolarisé + Rural				
	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans		
moto125	126 716	128 891	196 197	298 101	280 016	496 776	70 135	79 842	198 159	1 874 833	31%
moto125-750	86 332	69 143	164 358	205 935	245 453	520 914	67 955	99 196	230 356	1 689 642	28%
moto>750	47 255	48 125	77 871	146 805	172 897	250 296	48 316	67 798	110 835	970 198	16%
Cyclos	144 513	58 468	5 014	550 866	290 002	56 583	226 876	130 299	29 149	1 491 770	25%
	7%	5%	7%	20%	16%	22%	7%	6%	9%	6 026 443	100%

1.3. La taille de l'échantillon

Entre la base de sondage établie au 1^{er} janvier 2012 (6,03 millions d'observations ; tableau 1) et l'estimation fournie par l'ENTD (2,99 millions), le rapport est de 0,5. Ce rapport serait en réalité plus élevé, puisque le tirage a probabilités inégales (cf. infra) surreprésente les véhicules récents, plus susceptibles de rouler. D'un autre côté, la non-réponse diminue le taux de retour. Pour avoir environ 15 000 questionnaires exploitables, en appliquant ce taux et en tenant compte des « NPAI », la taille de l'échantillon est portée à 30 000 unités.

En appliquant les taux de sondage déterminés au § 1.4 ci-après, sur la base d'un échantillon de 30 000 véhicules, c'est en Corse et dans le Limousin que les allocations sont les plus faibles (environ 290 en Limousin, soit 150 retours exploitables). Il est donc possible de donner, y compris pour les régions les moins biens dotées, une estimation du parc par grands segments. Il n'est pas possible de fournir un niveau de détail fiable pour chacun des 5 DOM présents dans l'échantillon. En revanche, l'allocation moyenne pour l'ensemble des DOM est 2,5 fois plus élevée que dans le Limousin.

1.4. Allocation optimale de Neymann sous contrainte de coût

Pour le sondage, il est également tenu compte d'un plus fort taux de déchets parmi les véhicules âgés : « plus le véhicule est âgé dans le parc Sidiv, moins il a de chance d'être encore en service ». En effet, il serait coûteux d'envoyer le même « taux » de questionnaires aux véhicules très âgés, dont une plus grande proportion ne roule plus ou n'existe plus.

Une fonction de « coût » est alors introduite. Celle-ci a pour but de prendre en compte le surcoût d'interrogation de véhicules dans certaines strates lié à la surreprésentation présumée (et statistiquement vérifiée avec les résultats de l'ENTD) de véhicules présents à tort dans la base de sondage. Elle conduit à augmenter le taux moyen de retour de questionnaires remplis.

Une allocation optimale de Neymann est donc réalisée, en prenant en compte cette fonction qui représente le coût relatif d'obtention d'un questionnaire rempli pour un questionnaire envoyé. Cette fonction n'étant pas calculable *a priori*, elle est estimée, pour chacune des strates constituées, par le rapport entre le nombre de véhicules dans la base de sondage et celui des véhicules roulants estimé à partir de l'ENTD 2008, source auxiliaire de référence utilisée dans le plan de sondage.

$$C_h = \frac{N_h^{Carte\ grise}}{N_h^{ENTD}} = \text{fonction de « coût » unitaire, dans la strate } h / h \in [1, \dots, H = 36]$$

avec :

$$N_h^{Carte\ grise} = \text{effectif de la base des cartes grises, pour la strate } h$$

$$N_h^{ENTD} = \text{effectif (pondéré) de la base ENTD 2008, pour la strate } h$$

$$C = \sum_{h=1}^H n_h \cdot C_h = \text{« coût » global}$$

$$n = \sum_{h=1}^H n_h = \text{taille de l'échantillon (ici, } n = 30\,000)$$

Ce coût C_h augmente effectivement avec l'âge du véhicule, sauf pour les cyclomoteurs, en raison de la structure de la base des cartes grises : les cyclomoteurs n'y ont été introduits qu'en 2004, et la base s'enrichit progressivement avec la nouvelle loi obligeant l'immatriculation des cyclomoteurs circulants d'avant 2004.

Fonction de coût C_h (ici, pour un des trois zonages retenus) :

	Aire urbaine des grands pôles		
	moins de 5 ans	de 5 à 10 ans	de 11 à 29 ans
Moto 125 cm ³	1,7	2,1	6,2
Moto 125-750 cm ³	1,8	1,7	5,6
Moto > 750 cm ³	1,8	2,0	4,1
Cyclomoteurs	1,2	4,2	1,1

L'allocation optimale de Neyman sous cette contrainte de coût C_h est alors donnée par :

$$n_h = a_h \cdot \frac{C}{\sum_{h=1}^H N_h^{Carte\ grise} \cdot S_h^{ENTD} \cdot \sqrt{C_h}}$$

avec :

$$a_h = \frac{N_h^{Carte\ grise} \cdot S_h^{ENTD}}{(C_h)^{1/2}}$$

et : S_h^{ENTD} désignant l'écart-type de la strate h , issu de l'ENTD pour le kilométrage annuel moyen (variable directrice de l'allocation)

Sur la base de 30 000 véhicules sondés, ce qui représente un taux de sondage moyen de 0,50 %, les allocations de Neymann calculées à partir des données issues de la base de sondage des cartes grises d'une part, et des calculs effectués à partir de l'ENTD d'autre part, aboutit alors aux allocations suivantes :

Tableau 2

Allocation optimale de Neymann (n_h) - Échantillon de 30 000 2RM

n_h	Aire urbaine de Paris			Aire urbaine des grands pôles			Petite et moyenne aire urbaine + Multipolarisé + Rural			Sous-totaux	
	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans		
moto125	917	795	302	2 868	2 094	1 138	122	168	1 395	9 799	33%
moto125-750	1 156	516	313	1 192	1 430	1 433	401	263	352	7 056	24%
moto>750	498	366	279	1 311	911	712	335	315	913	5 640	19%
Cyclos	398	64	4	4 093	426	116	2 020	358	26	7 505	25%
	TOTAL (Taille de l'échantillon) :									30 000	100%
Sous-totaux :	2 969	1 741	898	9 464	4 861	3 399	2 878	1 104	2 686	30 000	100%
	10%	6%	3%	32%	16%	11%	10%	4%	9%		

Pour les seuls cyclomoteurs, les allocations sont faibles parmi les véhicules de plus de 11 ans : il est donc décidé de regrouper les véhicules âgés de 5 ans ou plus, sauf pour les aires urbaines des grands pôles (H = 34 désormais).

Au final, les taux de sondage varient de 0,11 % à 1,34 % selon les 34 strates constituées :

Taux de sondage dans chaque strate (en %)

f_h	Aire urbaine de Paris			Aire urbaine des grands pôles			Petite et moyenne aire urbaine + Multipolarisé + Rural		
	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans	moins de 5 ans	de 5 à 11 ans	Plus de 11 ans
moto125	0,72	0,62	0,15	0,96	0,75	0,23	0,17	0,21	0,70
moto125-750	1,34	0,75	0,19	0,58	0,58	0,28	0,59	0,27	0,15
moto>750	1,05	0,76	0,36	0,89	0,53	0,28	0,69	0,46	0,82
Cyclos	0,28	0,11		0,74	0,15	0,21	0,89	0,24	

Au final, les taux de sondage déterminés par l'allocation sont plus réduits pour les motocyclettes très âgées qu'ils auraient pu l'être sans intégrer la fonction de coût C_h , et inversement pour les véhicules plus récents.

2. Cyclomoteurs : sondage stratifié à probabilités égales

Un sondage stratifié à probabilités égales est effectué parmi les cyclomoteurs (7 strates).

Il faut noter que du fait de l'immatriculation plus récente des cyclomoteurs, la base de sondage ne permet pas toujours d'identifier les « vieux » cyclomoteurs. Dans la base Sidiv des cartes grises, ces « vieux » cyclomoteurs ont une date de 1^{ère} mise en circulation récente, qui ne correspond pas à la réalité (ils ont circulé avant que leur immatriculation ne devienne obligatoire). C'est pourquoi, dans les traitements d'après collecte, et pour les seuls cyclomoteurs, lorsque la date d'achat (obtenue via l'enquête) était inférieure à la

date de 1^{ère} mise en circulation (connue dans la base des cartes grises), le calcul de l'âge du véhicule a été estimé à l'aide de la date d'achat et de non de la date de 1^{ère} mise en circulation renseignée dans Sidiv.

3. Motocyclettes : sondage stratifié à probabilités inégales après prise en compte d'une information auxiliaire présente dans la base de sondage

En plus de la date de 1^{ère} mise en circulation – qui permet de déterminer l'âge du véhicule –, la base des cartes grises (Sidiv) comporte éventuellement une date postérieure relative à un changement de titulaire, d'adresse, d'état civil, ... ou bien encore une date relative à une réimmatriculation du véhicule (véhicule d'occasion, première immatriculation en série « normale », ...), ou à une autre déclaration, etc. Cette information ne concerne presque exclusivement que les motocyclettes, dans la mesure où les cyclomoteurs n'ont été intégrés qu'assez récemment dans la base. Cette date, qui confirme l'existence récente du véhicule, permet de distinguer des véhicules « confirmés » et des véhicules « non confirmés ».

Il est donc possible de déterminer, notamment pour les strates correspondant aux véhicules de plus de 5 ans, un sous-groupe comportant des véhicules plus « susceptibles de rouler » que d'autres (véhicules « confirmés »). Et donc de diminuer le taux de « déchets » lors de la collecte. Cette information auxiliaire, non présente dans l'ENTD, permet d'affiner l'allocation de Neymann initiale.

Ainsi, l'existence de 42 % des motocyclettes âgées de 5 à 11 ans est confirmée dans le parc des cartes grises après le 1^{er} janvier 2008 ; de même, l'existence de 21 % des motocyclettes de plus de 11 ans est confirmée dans le parc des cartes grises après le 1^{er} janvier 2008.

Dans chacune des 18 strates relatives aux motos de plus de 5 ans, un sondage à probabilités inégales est alors effectué, en tenant compte du fait que l'existence du véhicule soit ou non confirmée après 2008. Sont alors créées 36 « sous-strates » à la place des 18 strates initiales. La probabilité de tirage dans chaque sous-strate des motocyclettes dont l'existence est confirmée après le 1^{er} janvier 2008 est plus élevée, en tenant compte du rapport [nombre de motos confirmées / nombre de motos non confirmées] calculé dans chaque strate. Entre les strates qui comportent des véhicules confirmés et celles qui n'en comportent pas, le rapport des poids de sondage varie de 1,5 à 3,0 selon les strates (tableau 3).

4. Plan de sondage final

Le plan de sondage définitif comporte alors 52 strates, et les taux de sondage varient de 0,11 % à 1,36 %. Les poids de sondage varient de 73 à 935 (tableau 3).

Tableau 3a

Échantillon (véhicules de moins de 30 ans)

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Total
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C	
Cyclo	398	45	20	0	3	4093	241	185	3	113	2020	149	174	0	61	7 505
Moto 125	917	330	465	220	82	2868	887	1207	714	424	122	64	104	789	606	9 799
Moto 125-750	1156	189	327	205	108	1192	608	822	768	665	401	91	172	182	170	7 056
Moto >750	498	142	224	186	93	1311	351	560	367	345	335	126	189	435	478	5 640
																30 000

Tableau 3b

Poids de sondage dans chacune des 52 strates (arrondis à l'unité)

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé				
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans	
		NON CONF.	CONFIRMÉ	NON CONF.	CONFIRMÉ		NON CONF.	CONFIRMÉ	NON CONF.	CONFIRMÉ		NON CONF.	CONFIRMÉ	NON CONF.	CONFIRMÉ
Cyclo	363	934				135	681				488	415			
Moto 125	138	235	107	784	261	104	191	91	577	206	575	671	335	196	73
Moto 125-750	75	186	98	671	231	173	215	143	511	196	169	476	317	935	374
Moto >750	95	188	90	367	131	112	252	148	500	200	144	279	174	175	73

Annexe 2

Traitement de la non-réponse totale

À l'issue de la collecte, le taux de réponse à l'enquête s'est élevé à 49,6 %, le taux de NPAI (N'habite pas à l'adresse indiquée) a été de 13,8 %, et le taux de non-répondants de 36,6 % (tableau 4).

Tableau 4

Répartition, à l'issue de la collecte, selon les 52 strates :

- des effectifs de l'échantillon
- du nombre de répondants
- du nombre de NPAI
- du nombre de non-répondants

C = 2RM confirmé dans Sidiv
NC = 2RM non-confirmé dans Sidiv

4.1 - Échantillon (véhicules de moins de 30 ans)

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Total
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C	
Cyclo	398	45	20	0	3	4093	241	185	3	113	2020	149	174	0	61	7 505
Moto 125	917	330	465	220	82	2868	887	1207	714	424	122	64	104	789	606	9 799
Moto 125-750	1156	189	327	205	108	1192	608	822	768	665	401	91	172	182	170	7 056
Moto >750	498	142	224	186	93	1311	351	560	367	345	335	126	189	435	478	5 640
																30 000

4.2 - Répondants

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Total
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C	
Cyclo	154	14	10	0	1	1783	82	86	0	59	1023	60	95	0	31	3 398
Moto 125	462	131	204	46	38	1548	399	620	259	208	68	38	55	324	327	4 727
Moto 125-750	636	84	178	57	55	645	269	463	291	343	255	51	105	76	94	3 602
Moto >750	295	56	144	52	52	811	188	357	142	188	205	65	123	212	270	3 160
																14 887

4.3 - NPAI

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Total
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C	
Cyclo	52	13	0	0	0	560	73	16	0	4	200	31	14	0	1	964
Moto 125	111	83	57	112	11	313	198	142	247	57	9	11	12	196	54	1 613
Moto 125-750	96	47	24	98	12	100	125	74	222	73	23	13	12	47	19	985
Moto >750	24	29	8	81	4	77	60	25	105	22	10	18	10	80	31	584
																4 146

4.4 - Non Répondants

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Total
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C	
Cyclo	192	18	10	0	2	1750	86	83	3	50	797	58	65	0	29	3 143
Moto 125	344	116	204	62	33	1007	290	445	208	159	45	15	37	269	225	3 459
Moto 125-750	424	58	125	50	41	447	214	285	255	249	123	27	55	59	57	2 469
Moto >750	179	57	72	53	37	423	103	178	120	135	120	43	56	143	177	1 896
																10 967

Les taux de réponse varient de 21 % à 65 % selon les strates (tableau 5). Ils sont, comme attendus, supérieurs au sein des strates comportant des véhicules « confirmés », en comparaison des taux obtenus parmi les véhicules dont l'existence n'a pas été confirmée dans Sidiv après 2008. Ils sont également croissants avec la cylindrée du véhicule.

Tableau 5

Taux de réponse par strates

Taux de réponse (en %)

	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Total
	< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		< 5 ans	De 5 à 11 ans		Plus de 11 ans		
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C	
Cyclo	39	37				44	39				51	48				45
Moto 125	50	40	44	21	46	54	45	51	36	49	56	59	53	41	54	48
Moto 125-750	55	44	54	28	51	54	44	56	38	52	64	56	61	42	55	51
Moto >750	59	39	64	28	56	62	54	64	39	54	61	52	65	49	56	56
																50

1. Méthode de redressement de la non-réponse totale : modèle de réponse (MDR)

Pour le traitement de la non-réponse totale de l'enquête 2RM, le principal problème est de ne pas disposer de données externes fiables pour déterminer des marges de calage. L'ENTD 2008 a fourni une estimation du parc circulant de 2RM, mais la présente enquête a également pour but de déterminer un niveau de ce parc, et il n'est donc pas possible de se caler dessus. D'autre part, le fichier des cartes grises (SIDIV), à partir duquel est extrait la base de sondage de 6 millions de véhicules, contient beaucoup plus de véhicules que la réalité. Une certaine partie d'entre eux ne roule plus ou n'existe plus, sans qu'il soit possible de déterminer lesquels. Un calage sur marges est donc impossible à mener.

L'examen des réponses au questionnaire de l'enquête permet d'identifier 2 types de véhicules :

- les véhicules faisant partie du parc à la disposition des ménages (dénommés par la suite « **véhicules du parc** ») :

il s'agit de véhicules encore en possession du titulaire de la carte grise au 1^{er} mars 2012, qui sont en état de marche ou qui sont en panne mais qui devraient être réparés ; il faut noter que ces véhicules ne circulent pas forcément au cours des 12 derniers mois (leur kilométrage annuel peut être nul, ce qui permettra d'isoler ultérieurement un « parc non circulant » lors de l'exploitation) ;

- les véhicules ne faisant pas partie du parc à la disposition des ménages (dénommés par la suite « **véhicules hors parc** ») :

il s'agit de véhicules encore en possession du titulaire de la carte grise au 1^{er} mars 2012, mais qui sont hors d'usage et qui ne seront plus réparés, ou bien de véhicules qui ne sont plus en possession du titulaire de la carte grise au moment de l'enquête (véhicule détruit suite à un accident, mis à la casse, volé, abandonné, revendu, ...).

L'enquête visant à établir un parc à la disposition des ménages, puis en déduire un parc circulant, les informations sur les véhicules hors parc ne sont pas prises en compte ; généralement, le répondant n'a de toutes façons renseigné que le tout début du questionnaire, et il n'est pas invité à renseigner les caractéristiques du véhicule qu'il ne possède plus. En revanche, leur nombre va permettre de redresser les poids statistiques des unités restantes.

La méthode de redressement de la non-réponse totale s'appuie en effet sur les hypothèses suivantes (modèle de réponse) :

- Hypothèse H1 : au sein des véhicules confirmés, parmi les non-répondants, le taux de véhicules hors parc est identique à celui calculé parmi les répondants (cf. transcription au § 1.1.).
- Hypothèse H2 : parmi les véhicules non confirmés, le taux de non-réponse est le même que celui calculé au sein des véhicules confirmés (cf. transcription au § 1.2.).
- Hypothèse H3 : les questionnaires revenus en « NPAI » correspondent en bonne partie à des 2RM qui n'existent plus (hors parc). Les personnes qui déménagent ont juridiquement un mois pour signaler leur changement d'adresse ; en pratique, ce temps est parfois dépassé. D'après des estimations menées sur d'autres genres de véhicules (Véhicules utilitaires légers en particulier, où l'échantillon est constitué de cartes grises dont le véhicule a été récemment contrôlé), les NPAI représentent au maximum 5 % des véhicules. Il est décidé de retenir pour l'estimation du parc un pourcentage n'excédant pas 5 % de NPAI ne correspondant pas à des véhicules présents à tort dans la base de sondage (donc correspondant à des véhicules supposés exister encore).

1.1. Transcription dans les strates de 2RM dont l'existence est confirmée dans Sidiv après 2008

Soit, pour une strate de véhicules confirmés (la mention ^C désigne une strate de véhicules confirmés) :

t^C = taux de sondage dans la strate

Rep_P^C = effectifs des répondants faisant partie du parc

Rep_HP^C = effectifs des répondants hors parc

Rep^C = effectifs totaux des répondants

$Rep^C = Rep_P^C + Rep_HP^C$

$Taux^C$ = taux de 2RM du parc parmi les répondants

$Taux^C = Rep_P^C / Rep^C$

$NoRep^C$ = effectifs des non-répondants dans la strate

$NPAI^C$ = effectifs des « NPAI » dans la strate

N^C = effectif total de la strate

$N^C = Rep^C + NoRep^C + NPAI^C$

Des hypothèses H1 et H3 dérivent alors les calculs suivants du ratio de calage R^C :

$$R^C = \frac{\text{Rep_P}^C + \text{Taux}^C \cdot (\text{NoRep}^C + 0,05 \cdot N^C)}{\text{Rep_P}^C} \quad \text{si } \text{NPAI}^C / N^C > 0,05$$

$$R^C = \frac{\text{Rep_P}^C + \text{Taux}^C \cdot (\text{NoRep}^C + \text{NPAI}^C)}{\text{Rep_P}^C} \quad \text{si } \text{NPAI}^C / N^C \leq 0,05$$

Le nombre de 2RM dans le parc vaut alors, dans chaque strate de véhicules confirmés :

$$\text{Parc}^C = \text{Rep_P}^C \times (1/t^C) \times R^C$$

$$\begin{aligned} \text{Le nouveau poids statistique du 2RM est alors : } \text{Poids}^C &= \text{Parc}^C / \text{Rep_P}^C \\ &= R^C / t^C \end{aligned}$$

Ne sont alors conservées dans la strate que les observations des véhicules faisant partie du parc (Rep_P^C observations) et leur poids statistique unitaire vaut alors Poids^C (au lieu de $1/t^C$ initialement). Toutes les autres observations (les Rep_HP^C véhicules hors parc) sont retirées de la strate.

1.2. Transcription dans les strates de 2RM dont l'existence n'est pas confirmée dans Sidiv après 2008

L'idée est de ne s'appuyer que sur les effectifs de véhicules non confirmés ayant répondu et faisant partie du parc, et de leur appliquer la même proportion de non-réponse (cf. hypothèse H2). En effet, le comportement de non-réponse global n'est sans doute pas le même dans les deux strates, et il est préférable d'appliquer le calage le plus robuste, calculé sur les véhicules confirmés. Ne sont donc pas pris en compte, parmi les véhicules non confirmés, les « NPAI » (NPAI^{NC}) et les non-répondants (NoRep^{NC}).

Ainsi, les effectifs des répondants (véhicules du parc) au sein de la strate des 2RM non confirmés sont multipliés par le ratio de calage R^C calculé initialement parmi les véhicules confirmés (et non par un ratio R^{NC} qui pourrait également être calculé, avec les seules données de la strate) :

$$\text{Parc}^{NC} = \text{Rep_P}^{NC} \times (1/t^{NC}) \times R^C \quad (\text{la mention }^{NC} \text{ désigne une strate de véhicules non confirmés})$$

$$\begin{aligned} \text{Poids}^{NC} &= \text{Parc}^{NC} / \text{Rep_P}^{NC} \\ &= R^C / t^{NC} \end{aligned}$$

Ne sont alors conservées dans la strate que les observations correspondant aux 2RM du parc (Rep_P^{NC} observations) et leur poids statistique unitaire est alors Poids^{NC} (au lieu de $1/t^{NC}$ avant le redressement). Toutes les autres observations (les Rep_HP^{NC} véhicules hors parc) sont retirées de la strate.

1.3. Estimation du nombre de 2RM à la disposition des ménages, puis du parc circulant

Pour les cyclomoteurs et pour les motos de moins de 5 ans, où tous les véhicules sont considérés comme « confirmés » dans le plan de sondage, le traitement § 1.1. ci-dessus s'applique également dans chaque strate.

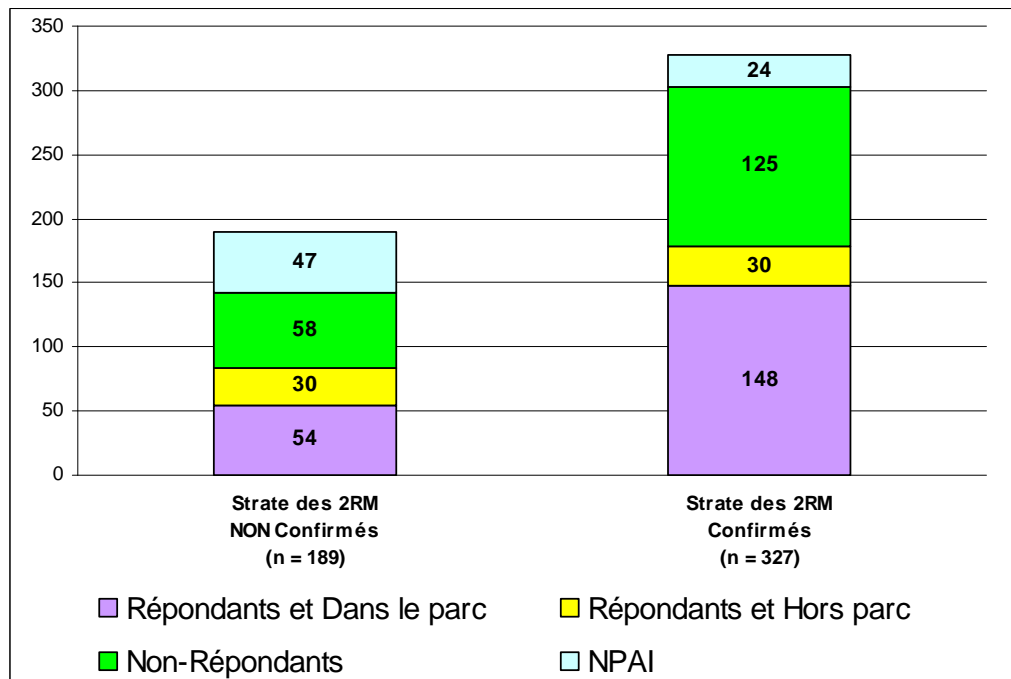
Une fois calculés dans les 52 strates, la somme des Parc^C et Parc^{NC} donne le niveau du parc de 2RM à la disposition des ménages. Dans ce parc, les véhicules qui n'ont pas roulé au cours des 12 derniers mois sont isolés (parc non circulant). Ceux qui ont parcouru au moins 1 kilomètre dans l'année constituent le parc circulant.

2. Exemple de calcul : motos de 125 à 750 cm³, âgées de 5 à 11 ans, situées dans l'aire urbaine de Paris

Dans le plan de sondage, les 516 motos de 125 à 750 cm³, âgées de 5 à 11 ans, et situées dans l'aire urbaine de Paris sont réparties dans deux strates, selon qu'elles soient ou non confirmées.

L'allocation des motos confirmées est de 327, et l'allocation de motos non confirmées est de 189 (Annexe 1, Tableau 3a). Les poids de sondage respectifs sont 98 (98,0228) et 186 (186,2434) (Annexe 1, Tableau 3b).

Après la collecte, la distribution des réponses dans les deux strates a été la suivante :



Taux de sondage : $t^{NC} \approx 1/186$ $t^C \approx 1/98$

2.1. Calcul du nombre de deux-roues dans la strate des 2RM confirmés

Parmi les $Rep^C = 178$ répondants, $Rep_P^C = 148$ deux-roues font partie du parc.

Le taux de 2RM dans le parc est donc : $Taux^C = 148/178 = 0,8315$

Ce taux est appliqué aux non-répondants $NoRep^C$: $125 \times 0,8315 = 103,9$ 2RM.

Le rapport entre les « NPAI » et l'effectif total de la strate est de $24/327 = 0,07$ (7 %). Ce taux étant supérieur à 5 %, il est retenu au plus 5 % de « NPAI » sur l'ensemble de la strate, soit $0,05 \times 327 = 16,35$.

Le taux $Taux^C$ est également appliqué à ce niveau : $16,35 \times 0,8315 = 13,6$ 2RM

Le « ratio de calage » vaut alors : $R^C = (148 + 103,9 + 13,6)/148 = 265,5/148$.

Le nombre de 2RM dans cette strate est de : $Parc^C = (265,5 * (1/t^C)) = 26\ 025$

Ne sont alors conservées dans la strate que les observations des véhicules faisant partie du parc (148 observations) et leur poids statistique unitaire est alors de $Poids^C = 26\ 025/148 = 175,8$ (au lieu de 98 initialement). Les autres observations ($Rep_HP^C = 30$ véhicules hors parc) sont retirées de la strate.

2.2. Calcul du nombre de deux-roues dans la strate des 2RM non confirmés

Les mêmes formules sont appliquées, mais le ratio de calage pris en compte demeure R^C :

$$\text{Parc}^{\text{NC}} = \text{Rep_P}^{\text{NC}} \times (1/t^{\text{NC}}) \times R^C = 54 \times (1 / [1/186,24]) \times (265,5/148) = 18\,046$$

$$\begin{aligned} \text{Poids}^{\text{NC}} &= \text{Parc}^{\text{NC}} / \text{Rep_P}^{\text{NC}} = 18\,046/54 = 334,1 \\ &= R^C / t^{\text{NC}} = (265,5/148) / (1/186,24) = 334,1 \end{aligned}$$

Ne sont alors conservées dans la strate que les 54 observations correspondant aux 2RM du parc, et leur poids statistique unitaire est alors $\text{Poids}^{\text{NC}} = 334,1$ (au lieu de 186 avant le redressement). Les 30 observations relatives aux véhicules hors parc sont retirées de la strate.

NB :

Dans cet exemple, le ratio R^{NC} , calculé à partir des effectifs de la strate des véhicules non confirmés en appliquant la méthode décrite au § 1.1. vaudrait $97,4/54$, soit 1,80. Il est pratiquement égal à R^C , qui vaut $265,5/148 = 1,79$.

3. Population des conducteurs principaux de 2RM

Le but de l'enquête est de déterminer le niveau du parc à la disposition des ménages et entreprises, puis d'en déduire un parc circulant et de décrire ses caractéristiques. De manière secondaire, il peut être intéressant d'estimer puis de caractériser la population des conducteurs principaux de 2RM.

Dans la base de véhicules qui constitue le parc à disposition, le total des poids véhicules donne le niveau du parc à disposition (et donc également le niveau du parc circulant) au 1^{er} mars 2012.

Il est également possible de calculer un poids conducteur, afin de constituer la population des conducteurs principaux de ces véhicules.

Ce calcul s'effectue parmi les répondants ayant renseigné que le véhicule était conduit par un seul conducteur, ou par un conducteur principal et un ou plusieurs conducteur(s) secondaire(s). Les véhicules conduits par plusieurs conducteurs, sans conducteur principal, ne sont pas pris en compte (question n° 2 du questionnaire), et leur poids conducteur est renseigné à valeur manquante.

Pour calculer le poids statistique relatif au conducteur de deux-roues, le poids véhicule est alors divisé par le nombre de 2RM dont dispose le conducteur (déterminé à la question n°30 du questionnaire), augmenté du véhicule échantillonné :

$$\text{Poids}^{\text{Conducteur}} = \text{Poids}^{2\text{RM}} / (N^{2\text{RM}} + 1)$$

où $N^{2\text{RM}}$ désigne le nombre de 2RM dont dispose le répondant (en plus du 2RM faisant l'objet de l'enquête).

4. Comparaison avec un redressement par Groupes homogènes de répondants (GHR)

La méthode des groupes homogènes de répondants (GHR) est une méthode de correction de la non-réponse totale par repondération.

4.1. Méthode des Groupes homogènes de répondants (GHR)

Étape 1 : modéliser une probabilité de réponse individuelle en fonction de l'information auxiliaire disponible

Dans l'enquête, l'essentiel de l'information auxiliaire disponible est résumé par deux variables :

- la variable « origine » décrit si le véhicule est issu ou non du parc du ministère de l'Intérieur, qui a servi à compléter la base Sidiv d'anciens véhicules pour constituer la base de sondage, après retrait des doublons présents dans les deux bases une fois la concaténation effectuée ;

- la variable « strate » synthétise l'information relative à l'âge, au type de cylindrée du véhicule et au type de zonage urbain (cf. Annexe 1, § 1.2.).

La première étape est donc une modélisation via une régression logistique du booléen « répondre à l'enquête », en fonction de ces deux variables. Les NPAI sont retirés de cette analyse (ils ne peuvent être intégrés avec les non-répondants). Dans le cadre de ce modèle, les 2 variables « origine » et « strate » sont significatives.

La régression permet alors d'estimer les probabilités de réponse de chaque unité statistique (option *pred=* dans la *proc logistic* de SAS). Ces probabilités pourraient être utilisées directement mais auraient tendance à surestimer le parc total des véhicules. En effet, une probabilité de réponse trop faible donnera un poids statistique trop important à l'individu statistique concerné (la probabilité de réponse estimée étant au dénominateur des nouveaux poids).

Étape 2 : constituer des groupes homogènes de réponse

La méthode utilisée est la méthode des quantiles égaux. Il est nécessaire d'ordonner les individus selon les probabilités estimées croissantes et de diviser l'ensemble des individus en groupes d'effectifs équivalents. En pratique, il est difficile d'obtenir des groupes de taille strictement équivalente, dans la mesure où beaucoup de probabilités estimées sont identiques. Dans cette analyse, 10 groupes de réponse ont été constitués, en utilisant les déciles fournis par la *proc univariate* de SAS (les résultats diffèrent peu avec 4 ou 5 groupes).

Étape 3 : estimation des probabilités de réponse

Dans chaque groupe constitué, une nouvelle probabilité de réponse est calculée, définie comme le rapport entre le nombre de répondants divisé par le nombre total d'observations. Cette probabilité de réponse est donc égale au taux de réponse à l'intérieur d'un groupe donné.

Étape 4 : calcul des nouveaux poids statistiques

Pour chaque répondant d'un GRH donné, les nouveaux poids statistiques sont définis comme le rapport entre les anciens poids statistiques et les probabilités estimées à l'étape 3. Les estimations du parc de 2RM s'effectuent alors en considérant ces nouveaux poids statistiques (est attribué aux non-répondants un poids nul).

4.2. Résultats et comparaisons avec le redressement « MDR »

Le parc à disposition est estimé à 4,04 millions de véhicules avec la méthode GHR (contre 4,14 millions par la méthode MDR - modèle de réponse, cf. § 1. & 2.). Le parc circulant de deux-roues, hors enduros et véhicules de 30 ans et plus, est estimé à 3,50 millions (tableau 6).

Tableau 6 - Parc circulant, hors enduros et véhicules de 30 ans ou plus (méthode GHR)

Redressement par "GHR"	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Ensemble	
	Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		(milliers)	(en %)
		NC	C	NC	C		NC	C	NC	C		NC	C				
Cyclo	92	4	11		2	370	40	91		46	160	21	52		18	908	25,9
Moto < 125	100	37	34	27	11	227	96	80	126	54	50	24	27	57	31	980	28,0
Moto 125-750	71	17	24	31	17	158	67	90	142	95	51	21	37	55	46	922	26,3
Moto > 750	41	18	18	28	10	126	62	71	90	57	45	24	28	45	29	693	19,8
Ensemble	304	76	86	88	39	880	265	333	405	206	306	91	145	175	105	3 503	100,0
Ensemble (en %)	8,7	2,2	2,5	2,5	1,1	25,1	7,6	9,5	11,6	5,9	8,7	2,6	4,1	5,0	3,0	100,0	

Ces résultats sont légèrement inférieurs à ceux obtenus par la méthode « MDR », décrite précédemment, mais ils n'incluent pas les NPAI. En intégrant ($Taux^C \cdot 0,05 \cdot N^C$, avec $Taux^C < 1$) de NPAI aux résultats, comme dans la méthode MDR (cf. § 1.1.) – ce qui revient à intégrer environ 3 % du parc à ce niveau, soit un peu plus de 0,1 million de véhicules supplémentaires, le niveau est quasiment égal à celui déterminé par la méthode « MDR » (tableau 7).

Tableau 7 - Parc circulant, hors enduros et véhicules de 30 ans ou plus (méthode MDR)

Redressement par "Modèle de réponse"	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Ensemble	
	Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		(milliers)	(en %)
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C		
Cyclo	97	4	11	2		391	38	86	48		170	22	54	19		942	26,0
Moto < 125	103	43	37	21	12	245	99	84	130	57	54	28	28	55	32	1 029	28,4
Moto 125-750	73	17	25	31	18	167	64	94	141	100	54	23	40	52	47	946	26,1
Moto > 750	43	13	19	23	10	136	62	75	89	60	45	22	29	46	30	703	19,4
Ensemble	317	78	92	77	40	939	263	339	407	216	323	96	152	173	109	3 621	100,0
Ensemble (en %)	8,7	2,2	2,5	2,1	1,1	25,9	7,3	9,4	11,2	6,0	8,9	2,6	4,2	4,8	3,0	100,0	

En structure, et par rapport à la méthode MDR, la méthode GHR conduit à très légèrement réduire le nombre de cyclomoteurs et de motos légères, tandis qu'elle accroît très légèrement le nombre de motos de plus de 125 cm³ (tableau 8).

Tableau 8 – Comparaison en structure

Structure GHR calée sur le total des effectifs Modèle de réponse	Aire urbaine de Paris					Aires urbaines des grands Pôles					Aires urbaines des petits/moyens Pôles + Rural + Multipolarisé					Ensemble		Écart
	Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		Moins de 5 ans	De 5 à 11 ans		Plus de 11 ans		(milliers)	(en %)	
		NC	C	NC	C		NC	C	NC	C		NC	C	NC	C			
Cyclo	95	4	11	2		382	42	94	47		166	22	54	19		939	25,9	-0,1
Moto < 125	103	38	35	28	12	234	99	83	130	56	52	25	28	59	32	1 013	28,0	-0,4
Moto 125-750	73	17	24	32	18	164	70	93	147	98	53	22	39	57	47	953	26,3	0,2
Moto > 750	43	19	19	29	10	130	64	74	93	59	46	25	29	46	30	716	19,8	0,3
Ensemble	314	79	89	91	40	910	274	344	418	213	316	94	150	181	109	3 621	100,0	0,0
Ensemble (en %)	8,7	2,2	2,5	2,5	1,1	25,1	7,6	9,5	11,6	5,9	8,7	2,6	4,1	5,0	3,0	100,0		
Écart (en points)	- 0,1	0,0	- 0,1	0,4	- 0,0	- 0,8	0,3	0,1	0,3	- 0,1	- 0,2	- 0,1	- 0,1	0,2	- 0,0	0,0		

Les résultats étant proches au final, la méthode MDR a été privilégiée. L'information auxiliaire mise en œuvre dans la régression qualitative est en effet assez pauvre (il n'y a pas assez d'information dans les variables explicatives pour prédire correctement la variable d'intérêt). De plus, le modèle initial de régression logistique n'est pas assez discriminant (les statistiques d'association ne sont pas suffisamment satisfaisantes pour retenir le modèle) : 54,4 % de paires concordantes (D de Sommers égal à 0,136, Gamma égal à 0,143). Enfin, le modèle ne prend pas en compte les NPAI. Toutefois, la méthode « GHR » a permis de « cadrer » l'estimation initiale fournie par la méthode MDR.

Annexe 3

Traitement de la non-réponse partielle

Les taux de non-réponse partielle sont assez faibles (notamment sur le kilométrage annuel) du fait d'un important travail qualitatif réalisé au cours de la collecte. Tout d'abord, de nombreux enquêtés ont contacté le responsable de l'enquête afin d'obtenir des précisions, limitant ainsi les problèmes d'incompréhension du questionnaire. Ensuite, de nombreux enquêtés ont également été contactés par téléphone ou courriel afin d'affiner leurs réponses ou de les corriger en cas d'incohérence manifeste, avant l'envoi à la saisie. Au final, plusieurs milliers de contacts téléphoniques ou par courriel ont permis de diminuer fortement les taux de non-réponse partielle.

Non-réponse partielle pour les variables qui ont été redressées

Variable	En %
Type de véhicule	0,9
Equipement d'un ABS	1,8
Motorisation du véhicule	1,6
Cycle du moteur essence (2 temps, 4 temps)	5,1
Assurance (au tiers, tous risques) du véhicule	1,5
Assurance contre vol et incendie	3,2
Assurance individuelle	6,4
Kilométrage annuel	1,1
Part kilométrage réalisée par le conducteur principal	4,9
Part du kilométrage réalisée en ville route et autoroute	3,0
Lieu de stationnement du véhicule au domicile	0,7
Période d'utilisation du véhicule	0,4
Fréquence d'utilisation du véhicule	0,6
Utilisation du véhicule la nuit	1,1
Utilisation du véhicule avec un passager	1,3
Fréquence d'utilisation du véhicule pour se rendre au travail	2,1
Utilisation des transports en commun pour se rendre au travail	2,5
Lieu de stationnement du véhicule sur le lieu de travail	4,1
Sexe du conducteur principal	1,0

Indicateur de précision

Coefficient de variation pour les estimations des kilomètres parcourus : 1,21 %

Traitement de la non-réponse partielle

Selon les questions, 3 méthodes d'imputations ont été utilisées afin de corriger la non-réponse partielle :

- méthode déductive
- méthode « hotdeck »
- méthode d'imputation par la moyenne

1. La méthode déductive

Cette méthode consiste à utiliser des relations logiques entre les variables.

Exemple d'utilisation : Question 8b

Si le véhicule est à motorisation essence, le moteur du véhicule est à :

- 2 temps
- 4 temps
- ne sais pas

Les manquants ont été redressés par la réponse la plus fréquente parmi les réponses connues des véhicules essence de même marque et de même modèle (la marque et le modèle sont des variables connues dans Sidiv et donc dans la base de sondage).

2. La méthode « hotdeck »

Dans le questionnaire, l'essentiel des variables est de type qualitatif. La donnée manquante est alors remplacée par une valeur observée pour un répondant choisi au hasard selon un sondage aléatoire simple avec remise.

Exemple d'utilisation : Question 18

Au cours de la période d'utilisation, ce véhicule est généralement utilisé :

- tous les jours ou presque
- 4 à 5 jours par semaine
- 1 à 3 jours par semaine
- seulement le week-end
- pratiquement jamais

La correction de la non-réponse a été réalisée sur le parc circulant, hors véhicules tout-terrain et véhicules de 30 ans et plus, par type de véhicule (cyclomoteurs, motocyclettes < 125 cm³, motocyclettes 125-750 cm³, motocyclettes > 750 cm³).

Résultats avant imputation (i.e. calcul sur les répondants), avec les poids de sondage :

	Cyclo (%)	Moto 125 (%)	Moto 125-750 (%)	Moto >750 (%)	Total Moto (%)	Ensemble (%)
Tous les jours ou presque	40,4	22,6	12,6	13,5	16,7	22,9
4 à 5 jours par semaine	18,5	17,8	15,2	11,2	15,2	16,0
1 à 3 jours par semaine	24,1	33,2	41,4	45,5	39,3	35,3
Seulement le Week-end	7,3	13,3	19,4	21,7	17,6	14,9
Pratiquement jamais	9,7	13,1	11,4	8,1	11,2	10,8
Ensemble	100,0	100,0	100,0	100,0	100,0	100,0

Résultats après imputation (i.e. calcul avec répondants + redressés), avec les poids de sondage :

	Cyclo (%)	Moto 125 (%)	Moto 125-750 (%)	Moto >750 (%)	Total Moto (%)	Ensemble (%)
Tous les jours ou presque	40,6	22,5	12,5	13,9	16,7	22,9
4 à 5 jours par semaine	18,4	17,8	15,3	11,2	15,2	16,0
1 à 3 jours par semaine	23,9	33,2	41,3	45,3	39,3	35,3
Seulement le Week-end	7,3	13,2	19,3	21,6	17,5	14,9
Pratiquement jamais	9,8	13,2	11,6	8,0	11,3	10,9
Ensemble	100,0	100,0	100,0	100,0	100,0	100,0

3. La méthode d'imputation par la moyenne

La donnée manquante, de nature quantitative, est remplacée par la moyenne calculée sur les répondants, pour chacune des 52 strates. Si les effectifs sont trop bas, ce qui imputerait un niveau éventuellement caricatural, des regroupements de strates *ad hoc* sont menés. Dans l'enquête, le nombre de répondants était suffisant dans chacune des strates où l'imputation a été menée.

Exemple d'utilisation : Question 13

Au cours des 12 derniers mois, la proportion des kilomètres parcourus par le véhicule est estimée à :

- ___ % en ville
- ___ % sur route
- ___ % sur autoroute

Si aucune des parts n'est renseignée, les parts redressées sont imputées par la moyenne des répondants concernés de la strate (véhicules circulants hors véhicules de type tout-terrain).

Vérification de l'impact de ce redressement : *En %*

	Part ville	Part route	Part autoroute
Avant redressement	42,8	50,5	6,7
Après redressement	43,0	50,5	6,5