



Apprendre des parodonnées pour améliorer les protocoles de collecte : l'exemple d'EpiCov

Géraldine Charrance¹, Paul Cochet¹, Aude Leduc², Marine Bondon¹, Thomas Merly-Alpa¹

¹ Ined, Service des Enquêtes et Sondages

² Insee, Direction des statistiques démographiques et sociales



Contexte et objectifs

Quand on conçoit une enquête, on est amené à se poser des questions sur le protocole de collecte :

- quels types de numéro de téléphone privilégier ?
- quel protocole de contact et de relance mettre en œuvre ?
- comment articuler au mieux les différents modes de collecte ?

Ce travail a pour but de donner des recommandations relatives aux protocoles de collecte.

Il s'inscrit dans un triple contexte :

- utilisation encore récente de Fidéli pour tirer les échantillons des enquêtes ;
- investiguer la manière d'exploiter des « parodonnées » ;
- profiter du caractère innovant du protocole de l'enquête EpiCov.



L'enquête EpiCov

- Grande enquête de la statistique publique (Inserm-Drees) qui s'est montée dans l'urgence pendant le premier confinement (vague 1) et qui est devenue un panel (vague 4 en préparation)
- Un très gros échantillon tiré dans Fidéli : 371 000 personnes de 15 ans ou plus
- découpé en vingt lots aléatoirement avec des protocoles distincts (vague 1)
 - Lots 1 à 4 : multimode Internet et téléphone
 - Lots 5 à 20 : monomode Internet
- Collecte durant un mois du 2 mai au 1^{er} juin 2020

Taux de réponse à la vague 1 : 38 %

lots multimodes : 47 %

lots monomodes : 35 %

Dans cette présentation :

- Paradoxes de la vague 1
- Concernant la métropole (350 000 personnes)



Sommaire

Recommandations portant sur...

- 1 ... la collecte par téléphone (CATI)
- 2 ... la collecte par Internet (CAWI)
- 3 ... la collecte multimode (CATI-CAWI)



Collecte par téléphone : questions/données

Questions :

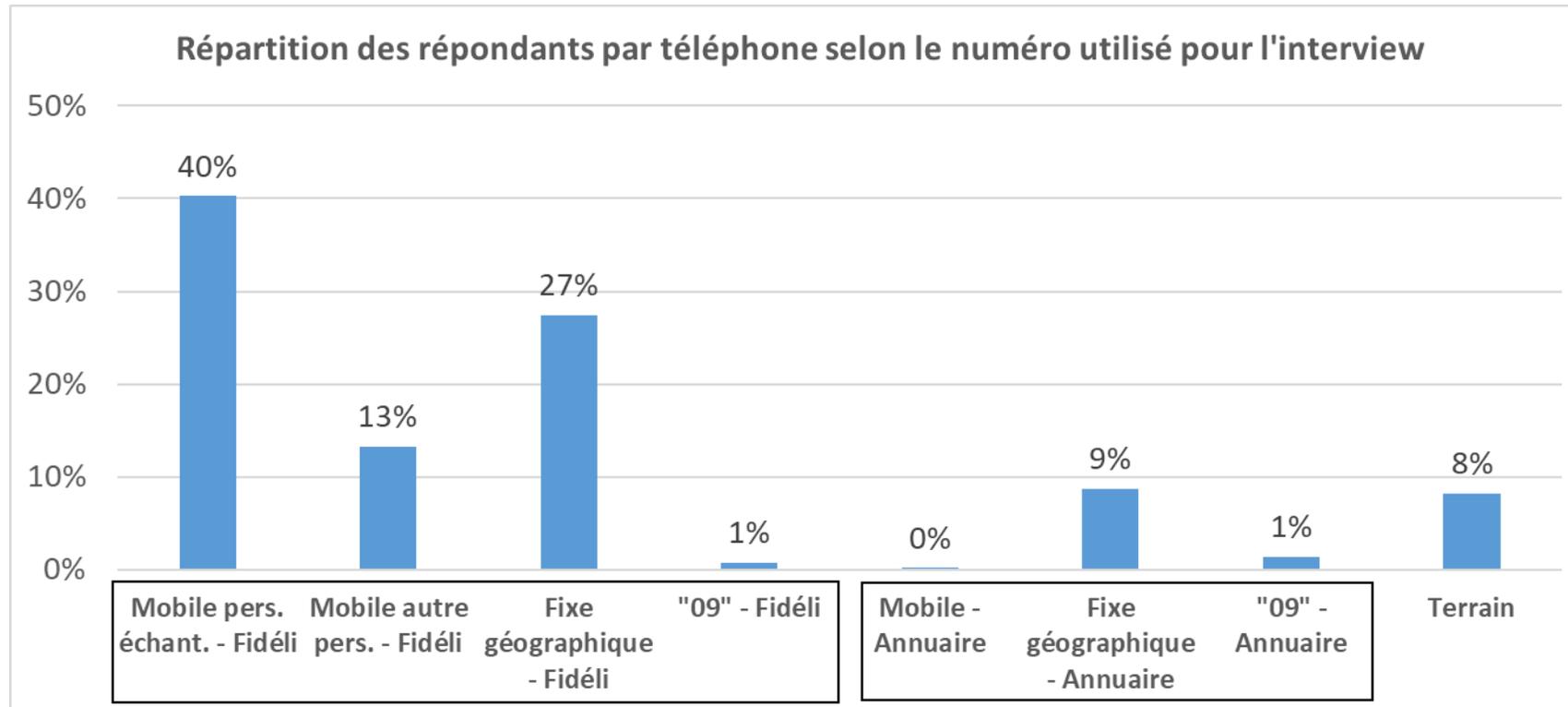
- Est-il utile de faire un enrichissement annuaire en plus des numéros Fidéli ?
- Quelle est l'utilité d'utiliser les numéros des « autres membres du ménage » dans une enquête « individus » ?
- Peut-on donner une priorisation des numéros à appeler ?

Données :

- Lots 1 à 3 avec enrichissement annuaire complet
- Historique des appels

| Total de 52 500 personnes | Part dans l'échantillon (en %) |
|--|--------------------------------|
| Au moins un numéro de téléphone issu de Fidéli | 72 |
| Mobile Fidéli de la personne échantillonnée | 31 |
| Mobile Fidéli d'une autre personne du ménage | 25 |
| Fixe géographique - Fidéli | 35 |
| « 09 » - Fidéli | 3 |
| Un numéro issu de l'enrichissement annuaire | 12 |
| Mobile - Annuaire | 0 |
| Fixe géographique - Annuaire | 10 |
| « 09 » - Annuaire | 2 |
| Au moins un numéro (Fidéli ou Annuaire) | 83 |

Nombre de questionnaires par téléphone selon le type de numéro



« Terrain » : numéros de téléphone récupérés en cours de collecte (hotline, enquêteurs, etc.)



Rentabilité des appels selon le numéro : « décroché » et interview

| | Probabilité qu'un appel aboutisse à un "décroché" | Probabilité que l'appel aboutisse à une interview sachant qu'il y a eu un "décroché" | Probabilité que l'appel aboutisse à une interview |
|-------------------------------|---|--|---|
| Constante | 16,9*** | 25,7*** | 4,3*** |
| Type de numéro | | | |
| Mobile pers. échant. - Fidéli | 4,4*** | 9,5*** | 3,2*** |
| Mobile autre pers. - Fidéli | 1,8*** | -3,6*** | |
| Fixe géographique - Fidéli | Réf. | Réf. | Réf. |
| "09" - Fidéli | -5,8*** | | -1,4** |
| Mobile Annuaire | -3,8** | | -1,7* |
| Fixe géographique Annuaire | -1,3*** | -5,5*** | -1,3*** |
| "09" Annuaire | -7,8*** | | -2,4*** |
| Numéro "terrain" | 12,0*** | 21,5*** | 9,8*** |

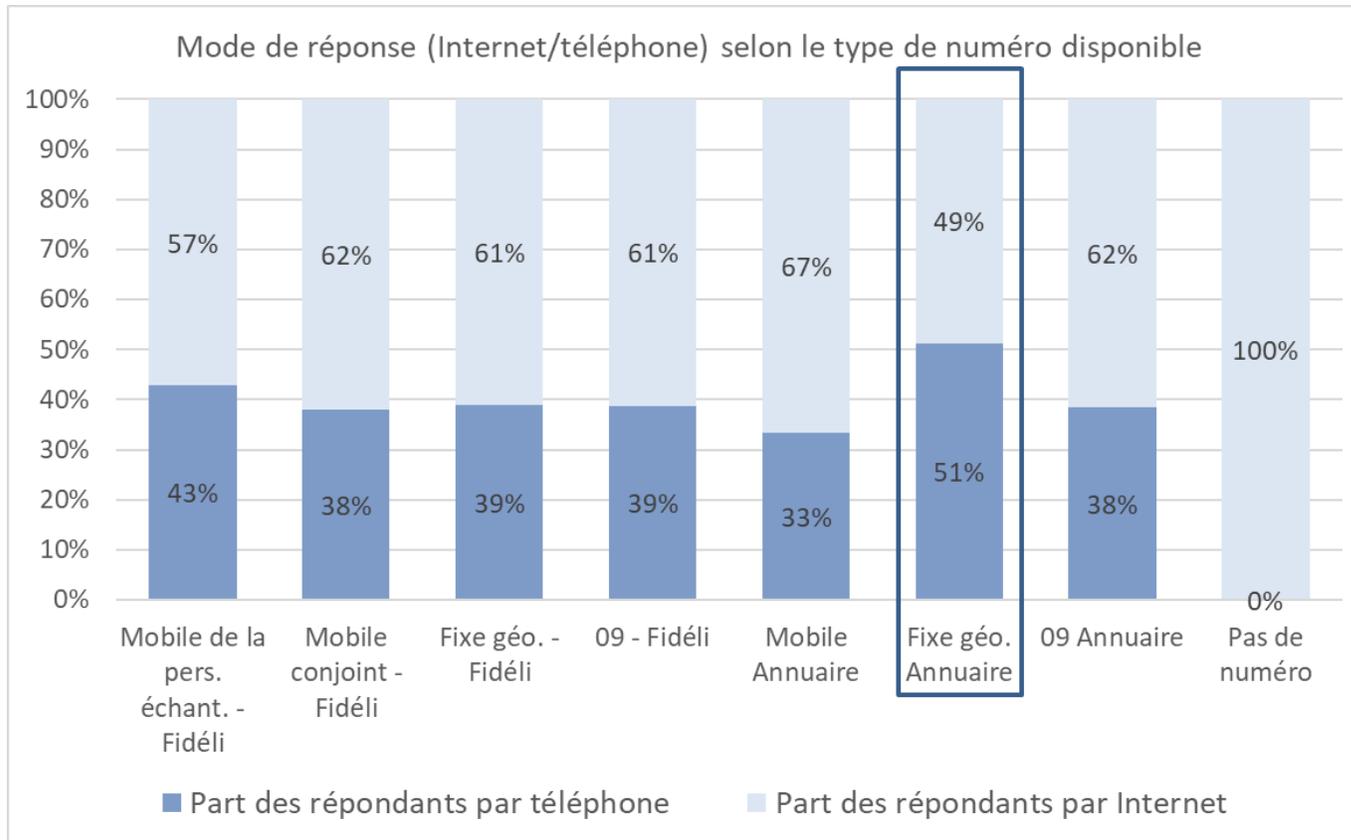
- Les appels sur les mobiles aboutissent plus fréquemment à un contact
- L'écart entre les personnes échantillonnées et les autres peut tenir à la lettre avis
- Résultats « Mobile » maintenus hors 1^{er} confinement ?
- Les Fixe annuaire aboutissent moins fréquemment à une interview que les Fixe Fidéli : qualité des numéros ?
- Les « 09 » aboutissent moins fréquemment à un contact : utiles ?
- Les téléphones « terrain » sont, comme attendu, les plus rentables

Variables de contrôle : sexe, âge en tranches décennales, QPV, décile de niveau de vie, pays de naissance, tranche d'unité urbaine, statut d'occupation du logement, moments des appels

*** : p-value < 0,001 ; ** : p-value < 0,01 ; * : p-value < 0,05



Avoir des profils de répondants différents avec l'enrichissement annuaire



- Réel apport de l'enrichissement téléphonique :
- Quantitatif et
 - en termes de réduction de biais de sélection

Caractéristiques des répondants :

| | Part parmi les répondants numéros Annuaire | Part parmi les répondants autres numéros |
|----------------|--|--|
| 65 ans ou plus | 48 % | 23 % |
| Déciles 1 à 4 | 45 % | 36 % |
| QPV | 7 % | 6 % |

- Évolution à l'avenir ?



Collecte par Internet : questions/données

Questions :

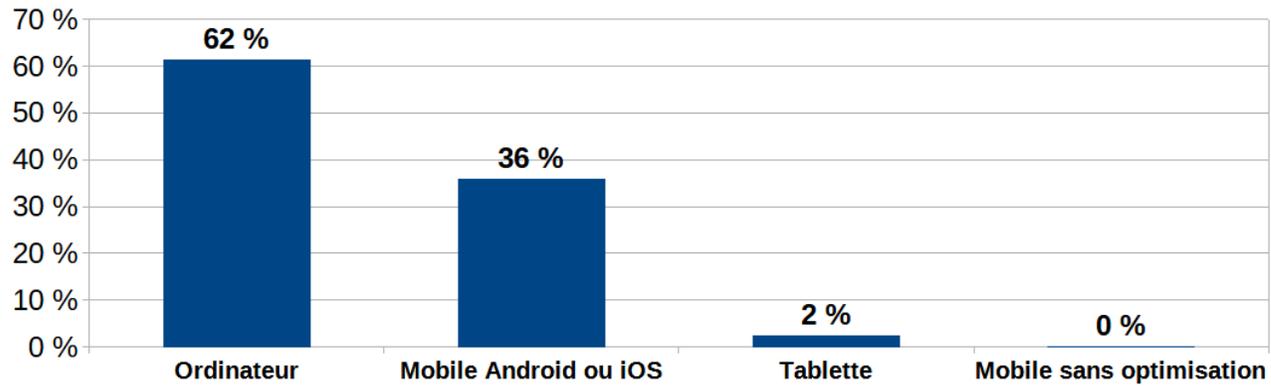
- Sur quel support les enquêtés répondent-ils ?
- Comment réduire les abandons par Internet ?
- Quel protocole de contact pour animer une collecte par Internet ?

Données :

- Données de gestion (suivi des contacts et informations de connexion Internet)
- Matériel et abandons étudiés à partir de l'ensemble des répondants par Internet : 115 000 personnes
- Protocole de contact : lots 5 à 20 avec une collecte exclusivement Internet

Collecte par Internet : plus d'un tiers répondent sur mobile

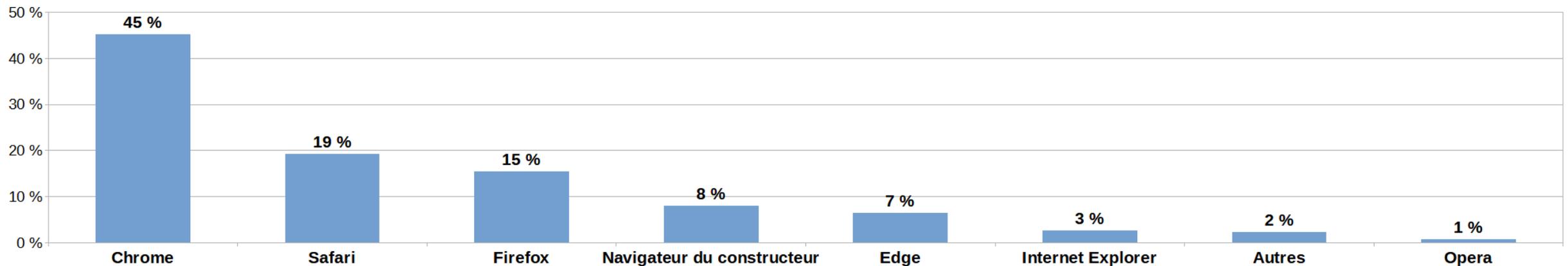
Répartition des répondants par Internet selon le support utilisé



Enjeu fort à ce que le questionnaire dans toute sa richesse fonctionne bien sur :

- mobile
- via les différents types de navigateur

Répartition des répondants par Internet selon le navigateur utilisé





Les abandons par Internet : plus marqués socialement

Internet :

6,4% des questionnaires commencés sont abandonnés en cours de collecte

Téléphone :

4,5 % des questionnaires commencés sont abandonnés en cours de collecte

Plus marqué socialement par Internet :

- Plus de coefficients significatifs dans la régression « Internet »

Outils/Conception orientés « mobile » à développer

| Matériel utilisé | |
|-------------------------|-------------|
| Mobile (Android ou iOS) | <i>Réf.</i> |
| Mobile (autres) | 20,4*** |
| Tablette | -3,2*** |
| Ordinateur | -3,9*** |

| | Téléphone | Internet |
|--------------------------|-------------|-------------|
| Constante | 3,0*** | 6,7*** |
| Sexe | | |
| Femmes | <i>Réf.</i> | <i>Réf.</i> |
| Hommes | 0,6* | 1,3*** |
| Inconnu | | 7,9*** |
| Âge | | |
| 15 - 24 ans | | |
| 25 - 34 ans | | 1,5*** |
| 35 - 44 ans | | 0,6* |
| 45 - 54 ans | <i>Réf.</i> | <i>Réf.</i> |
| 55 - 64 ans | | |
| 65 - 74 ans | 1,4* | |
| 75 - 84 ans | 2,6*** | 2,9*** |
| 85 ans et plus | 11,0*** | 13,6*** |
| Déciles de niveau de vie | | |
| 1er décile | | 2,4*** |
| 2ème décile | | 1,1** |
| 3ème décile | | 1,1* |
| 4ème décile | | 1,0* |
| 5ème décile | <i>Réf.</i> | <i>Réf.</i> |
| 6ème décile | | |
| 7ème décile | | |
| 8ème décile | | |
| 9ème décile | -1,1* | -1,4*** |
| 10ème décile | -1,2* | -1,4*** |
| Le ménage vit en QPV | | |
| Oui | | 1,5*** |
| Non | | <i>Réf.</i> |
| Lieu de naissance | | |
| France | <i>Réf.</i> | <i>Réf.</i> |
| Etranger | 2,5*** | 4,9*** |
| Inconnu | | |



Animer la collecte Internet : multiplier les contacts

| Coordonnées disponibles | Rattachement des coordonnées | Probabilité de répondre |
|-------------------------------------|------------------------------|-------------------------|
| Constante | | 27,9*** |
| Courrier distribué | | |
| | Personne échantillonnée | 26,1*** |
| + Mail | Autre personne du ménage | 16,1*** |
| | Personne échantillonnée | 18,1*** |
| + SMS | Autre personne du ménage | 17,2*** |
| | Personne échantillonnée | 37,7*** |
| + Mail et SMS | Autre personne du ménage | 26,4*** |
| Uniquement le courrier | | Réf. |
| Courrier non distribué (PND) | | |
| | Personne échantillonnée | -21,5*** |
| + Mail | Autre personne du ménage | -22,8*** |
| | Personne échantillonnée | -21,2*** |
| + SMS | Autre personne du ménage | -24,3*** |
| | Personne échantillonnée | -18,4*** |
| + Mail et SMS | Autre personne du ménage | -21,0*** |
| Aucune coordonnée | - | -27,2*** |

Variables de contrôle : sexe, âge en tranches décennales, QPV, décile de niveau de vie, pays de naissance, tranche d'unité urbaine, statut d'occupation du logement

*** : p-value < 0,001

Protocole de contact très chargé dans EpiCov : en un mois, au max : 1 courrier postal d'annonce + 1 courrier postal de relance + 3 SMS + 3 mails + 2 messages vocaux

Lots 5 à 20 : collecte Internet, quel effet de ces modes de contact ?

- Le courrier postal semble rester très important
- Multiplier les canaux de contact augmente la participation
- D'autant plus que les coordonnées sont celles de la personne échantillonnée
- Sauf pour les SMS (pas d'écart pers. échantillonnée et autre personne du ménage)

Mais possiblement de l'hétérogénéité inobservée



Collecte multimode : questions/données

Questions :

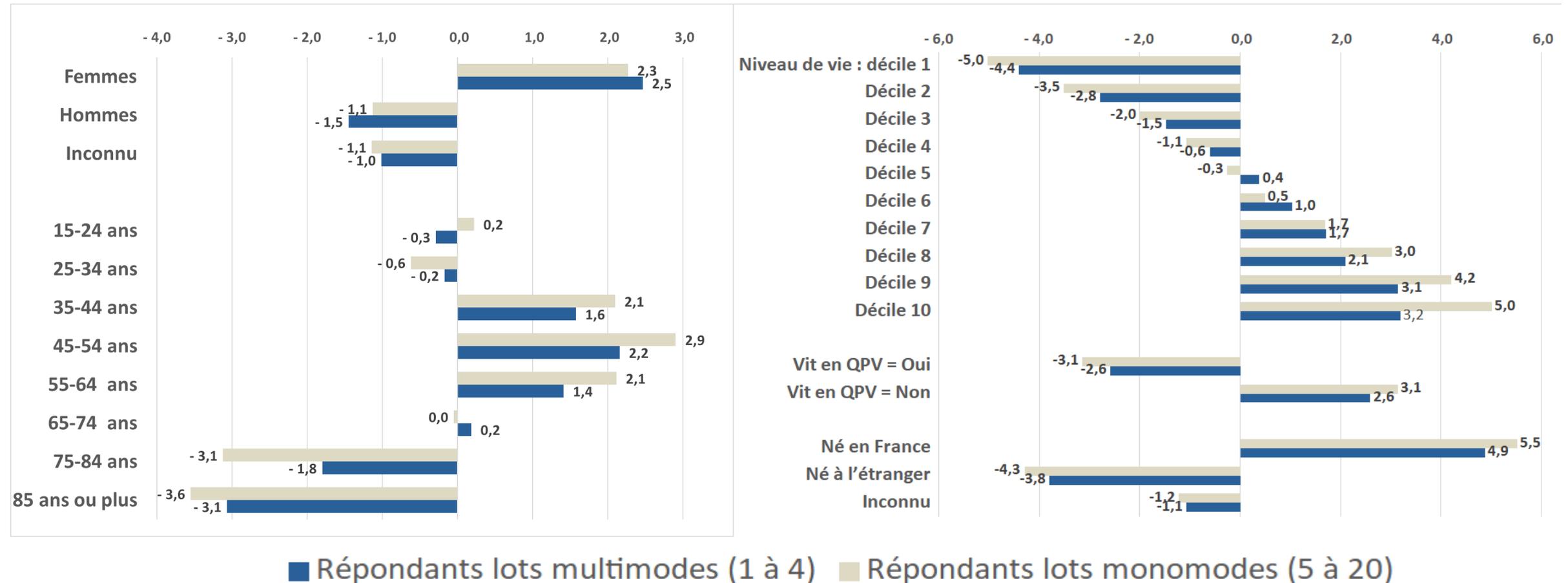
- Quel apport du multimode par rapport au monomode Internet ?
- Quel protocole multimode privilégier ?
- Quelles conséquences sur le profil des répondants ?
- et sur le coût de l'enquête ?

Données :

- Les lots 1-3 : multimode concurrentiel
- Le lot 4 : collecte d'abord Internet puis multimode (disponibilité enquêteurs)
- Rythme de collecte, taux de réponse, mode de réponse, nombre d'appels, etc.

Collecte multimode : moins de biais de sélection

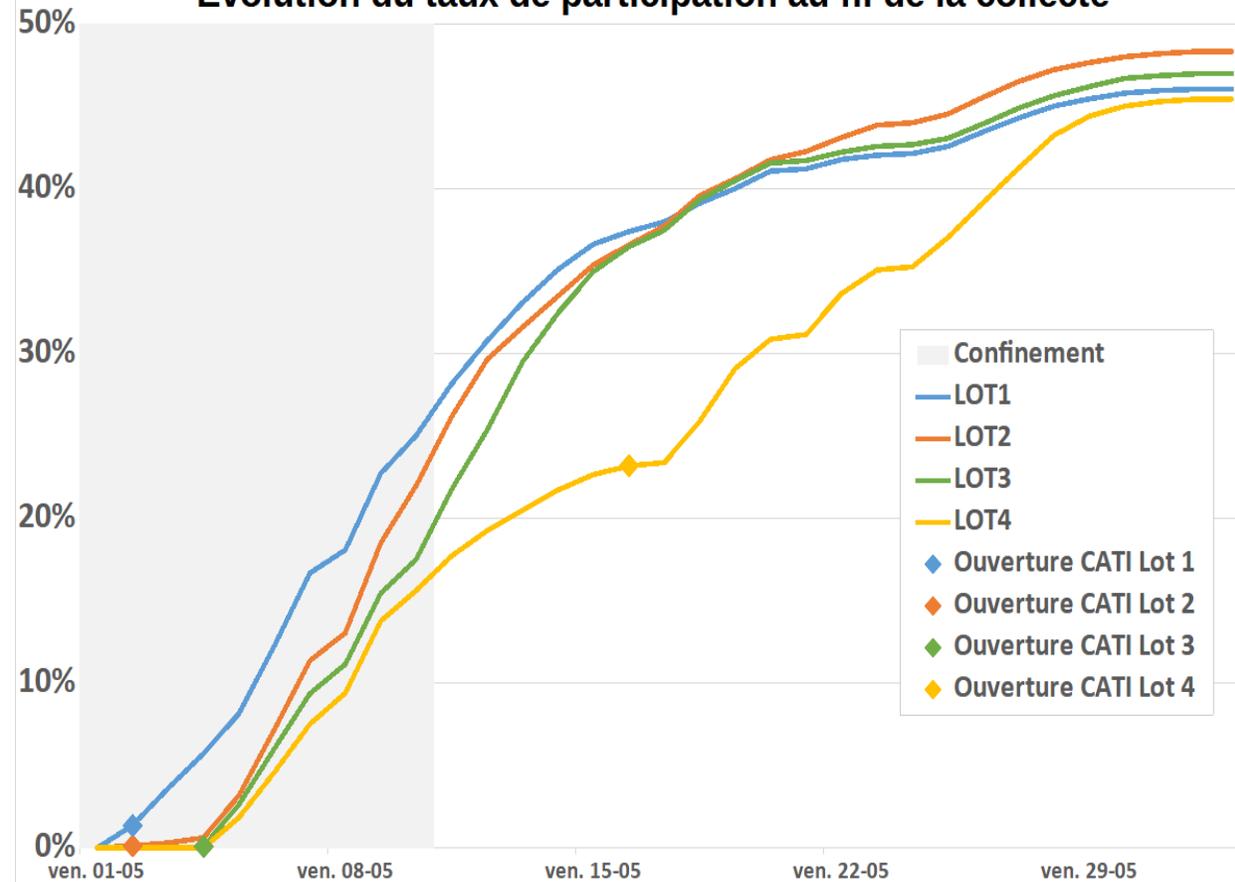
Écarts en points de pourcentage entre la structure des répondants et celle de l'échantillon initial





Ouverture de la collecte téléphone tardive : un taux de réponse comparable

Évolution du taux de participation au fil de la collecte



Ouverture tardive de la collecte par téléphone :

- Un taux de réponse comparable :
 lot 1 : 46 % ; lot 2 : 48 % ; lot 3 : 47 % ; lot 4 : 45 %
- Un rythme de collecte ralenti
- Une part de questionnaires par Internet plus importante quand la collecte téléphone est ouverte plus tard :

Part de questionnaires ...

| | Internet | téléphone |
|-------|----------|-----------|
| Lot 1 | 55 % | 42 % |
| Lot 2 | 61 % | 38 % |
| Lot 3 | 62 % | 36 % |
| Lot 4 | 74 % | 25 % |

Une collecte téléphonique plus légère et des profils de répondants similaires

Une collecte téléphonique plus légère :

- Moins de numéros appelés :

| | Nombre de numéros appelés | % des numéros appelés parmi les appelables |
|-------|---------------------------|--|
| Lot 1 | 13 714 | 94 % |
| Lot 2 | 13 653 | 94 % |
| Lot 3 | 12 414 | 85 % |
| Lot 4 | 9 917 | 71 % |

- Un coût plus faible : une économie d'environ un tiers du coût dans le lot 4 par rapport aux autres lots

Des profils de répondants similaires :

| | Taux de réponse | |
|----------------------|-----------------|------------|
| | Lots 1 à 3 | Lot 4 |
| Âge | | |
| 15 - 24 ans | 46 % | 44% |
| 25 - 34 ans | 45 % | 44% |
| 35 - 44 ans | 50 % | 49% |
| 45 - 54 ans | 52 % | 51% |
| 55 - 64 ans | 51 % | 50% |
| 65 - 74 ans | 49 % | 46% |
| 75 - 84 ans | 39 % | 36% |
| 85 ans ou plus | 21 % | 17% |
| Niveau de vie | | |
| Décile 1 | 32 % | 32% |
| Décile 2 | 35 % | 33% |
| Décile 3 | 39 % | 38% |
| Décile 4 | 45 % | 42% |
| Décile 5 | 49 % | 47% |
| Décile 6 | 52 % | 50% |
| Décile 7 | 56 % | 54% |
| Décile 8 | 58 % | 57% |
| Décile 9 | 63 % | 61% |
| Décile 10 | 63 % | 63% |
| Total | 47 % | 45% |



Conclusion : des exemples d'usage des parodonnées qui peuvent être prolongés

Mise en évidence de quelques usages qui peuvent être fait des parodonnées

- pour améliorer les connaissances sur les protocoles de collecte
- pour objectiver un certain nombre de savoirs que l'on a sur les protocole de collecte d'enquête

Prolongements possibles :

- Exploiter les parodonnées des DROM (taux de réponse par internet : 15 % versus 35 % pour la métropole)
- Investiguer l'aspect panel sur la constance des modes de réponse et l'attrition
- Profilage

Études à partir des données d'autres enquêtes en dehors du contexte Covid