

Répliquer une enquête face-à-face sur un access panel web ? Une comparaison multipanels.

Stéphane Legleye, Ined & Université Paris-Saclay, University of Paris-Sud, UVSQ, CESP, INSERM ; Sandra Hoibian, Crédoc ; Jérôme Cubillé, EDF ; Patricia Croutte, Crédoc, Géraldine Charrance, Ined

Résumé

Les panels d'internautes volontaires (access panels) permettent d'administrer rapidement des enquêtes à faible coût. Peut-on envisager d'y basculer des enquêtes traditionnelles ? Les processus de sélection des répondants diffèrent et des effets mode de collecte sont à anticiper pour tout ou partie des variables... Il faut donc expérimenter : comparer des redressements et éventuellement même, des panels entre eux.

Afin de gérer au mieux le passage en ligne de son enquête barométrique sur les, le CREDOC a répliqué la 37^{ème} vague de son enquête sur les Conditions de vie et les Aspirations, traditionnellement menée en face-à-face, sur un access panel.

La comparaison a été faite pour une batterie d'indicateurs en utilisant plusieurs méthodes de repondération (calage direct ou bien repondération par l'inverse du score de propension suivi d'un calage). Une des originalités est la prise en compte de variables auxiliaires censées approximer le satisficing et la désirabilité sociale, imputables à la présence/absence d'un enquêteur. Les résultats soulignent qu'aucun des neuf redressements ne permet de réduire totalement les écarts de mesure, et que deux méthodes doivent être écartées en raison de la trop grande dispersion des poids. La comparaison de sous-échantillons appariés sur des scores de propension incluant les variables auxiliaires confirme l'irréductibilité de la différence entre access panel et face-à-face pour quelques comportements d'usages de produits psychoactifs.

La comparaison des caractéristiques de trois panels commerciaux concurrents suggère, en attendant des analyses plus poussées à venir, que les résultats ont une certaine portée générale.

Mots clefs : multimode ; access panel ; Credoc ; satisficing ; désirabilité sociale ; score de propension

I. Introduction

Les enquêtes en population générale sont confrontées depuis de nombreuses années à une baisse significative de leur taux de participation, quel qu'en soit le mode et des efforts techniques et financiers de plus en plus importants sont nécessaires pour tenter d'enrayer ce phénomène (1). Parallèlement, les instituts de sondage ont, au moins pour les études marketing et certaines prestations à finalité non commerciales, abandonné l'échantillonnage traditionnel, même par quotas, pour se tourner vers le développement de panels internet (ou access panels), qui représentent le vivier d'enquêtés potentiels dans lequel ils effectuent directement leurs enquêtes. La popularité de ces enquêtes en ligne, leur coût et leur rapidité sont tels, que les enquêtes traditionnelles peinent de plus en plus à s'imposer auprès des financeurs et commanditaires non commerciaux. La question posée par ces panels est celle de la représentativité : les panélistes sont recrutés et enquêtés suivant un triple filtre. Ils sont d'abord équipés d'Internet, ce qui n'est le cas que de 75% de la population française (2-4) ; ensuite, ils ont des pratiques d'Internet qui les exposent à des sollicitations de recrutement pour un panel ; finalement, ils acceptent ensuite de participer ou non à une enquête du panel. De plus, ces panélistes sont rétribués à l'enquête (sous forme de bons cadeaux), ce qui jette un doute sur la sincérité de leurs déclarations et sur leurs motivations profondes.

Comparer les résultats collectés sur deux supports différents (internet et téléphone ou face-à-face) pose la question de la séparation des effets modes de collecte de l'effet de sélection des personnes ou tout au moins de leur réduction. Il est clair que les processus de sélection de personnes dans les panels internet commerciaux et les enquêtes en population générale diffèrent. De plus, l'absence d'enquêteur sur internet peut induire des réponses plus sincères ou plus extrêmes à des questions portant sur des thèmes sensibles (réduction de la désirabilité sociale), dès lors que certains comportements sont susceptibles d'être dévalorisés ou stigmatisants(5-7). Ce souci de donner une image positive de soi peut donc pénaliser les modes de collecte avec enquêteur, mais à l'inverse, l'absence d'enquêteur dans les enquêtes en ligne peut induire une dégradation de la qualité des réponses lorsque les questions sont complexes, nombreuses et répétitives ou que leur thème suscite peu l'intérêt de l'enquêté. Dans ce cas, l'attention de l'enquêté n'est pas stimulée par l'enquêteur, produisant un effet de non-maximisation ou « satisficing »¹ : le répondant se satisfait de réponses qu'il estime être le résultat d'efforts suffisants. Il a été proposé dans la littérature d'ajouter des questions attitudinales très générales dites à tort « webographiques » pour ajuster les comparaisons entre panels et enquêtes randomisées dans des analyses multivariées en tenant compte des différences de processus de sélection et traits comportementaux afférents (8). Il existe aussi des questionnaires mesurant la suffisaisance(9)et la désirabilité sociale (10, 11)que l'on peut envisager d'introduire dans les enquêtes afin de tester leur pertinence.

Pour faire suite à un précédent travail de recherche² et afin de gérer au mieux le passage en ligne de son enquête barométrique sur les Conditions de vie et les Aspirations, le CREDOC a organisé la 37^{ème} vague de son dispositif d'enquête de façon simultanée et identique, sur deux échantillons distincts, l'un interrogé en face-à-face, l'autre interrogé en ligne. La comparaison a été faite pour une batterie d'indicateurs en utilisant plusieurs méthodes de repondération en un ou deux temps (calage direct ou bien repondération par l'inverse du score de propension suivi d'un calage). Une des originalités de la comparaison est la prise en compte de variables censées approximer les comportements de réponse des individus et donc de neutraliser les tendances à la suffisaisance et à la désirabilité sociale, composantes théoriques majeures de la différence de mesure obtenue en présence ou en absence d'un enquêteur. Les résultats généraux soulignent que pour un ensemble assez large de comportements, aucune de ces méthodes ne permet de réduire les écarts de mesure entre les deux modes de collecte.

¹ Mot-valise contractant *satisficing* et *sufficing* dont on pourrait suggérer une traduction construite sur le même modèle par « suffisaisance »

²Régis Bigot, Patricia Crouette, Fanette Recours, Enquêtes en ligne : peut-on extrapoler les comportements et les opinions des internautes à la population générale ?, CREDOC, Cahier de recherche n°273, décembre 2010, <http://www.credoc.fr/pdf/Rech/C273.pdf>

La généralisabilité de ces résultats est questionnée à travers la comparaison des caractéristiques de plusieurs panels commerciaux concurrents.

Contexte

Pour étudier les effets d'un passage en ligne d'une enquête en population générale jusque-là réalisée en face à face, le CREDOC a organisé, de façon simultanée, deux interrogations identiques sur deux échantillons distincts, l'un interrogé en face-à-face, l'autre collecté en ligne via un panel d'internautes à la même période (décembre 2014 – janvier 2015). Cette opération s'est déroulée dans le cadre de l'enquête récurrente du CREDOC sur les Conditions de vie et les Aspirations des Français, enquête multithématique menée depuis 37 ans, en face-à-face, au domicile des personnes interrogées qui comporte à la fois un corpus de questions généralistes et des questions insérées à la demande de différents clients, le plus souvent publics.

Questionnaire

Le questionnaire – commun aux deux enquêtes – comporte un tronc commun de questions généralistes avec à la fois un descriptif sociodémographique étoffé et de nombreuses questions d'opinions (sur la famille, le moral économique, la confiance dans les différents acteurs, le traitement de la pauvreté, les inquiétudes, etc.). Ces questions généralistes sont financées par le CREDOC. S'y ajoutent des questions insérées par les différents clients qui souscrivent au système d'enquête. La plupart de ces questions étant confidentielles, les résultats et comparaisons entre face-à-face et online ne seront pas repris aussi.

Le questionnaire a été saisi à la fois en CAPI (ComputedAssistedPersonalInterviewing) et en CAWI (ComputedAssisted Web Interviewing) : le CREDOC s'est assuré de la totale harmonisation des deux questionnements. Si, par exemple, une liste de modalités de réponse était fournie aux enquêtés en face-à-face pour une question, toutes les modalités apparaissaient à l'écran pour l'internaute lors de la passation de cette même question. Si, à l'inverse, une batterie de questions était posée au fur et à mesure, ces mêmes questions apparaissaient sur des écrans différents pour l'internaute afin qu'il ne puisse pas prendre connaissance de l'ensemble des questions qui allaient lui être soumises sur ce thème.

L'enquête face-à-face

L'échantillon de 2 000 interviews est représentatif des individus habitant en France métropolitaine (hors Corse) âgés de 18 ans et plus, sélectionnés selon la méthode des quotas :

- **ZEAT (en 9 modalités)**
- **Taille del'unité urbaine** (en 9 modalités : moins de 2 000 habitants - de 2 000 à 5 000 habitants - de 5 000 à 10 000 habitants - de 10 000 à 20 000 habitants - de 20 000 à 50 000 habitants - de 50 000 à 100 000 habitants - de 100 000 à 200 000 habitants - plus de 200 000 habitants - Paris et agglomération)
- **Sexe** de l'interviewé,
- **Age** de l'interviewé en 5 modalités (18 à 29 ans - 30 à 49 ans - 50 à 59 ans - 60 à 69 ans - 70 ans et plus),
- **PCS** de l'interviewé en 8 modalités (travailleur indépendant - cadre et profession libérale - profession intermédiaire - employé - ouvrier - étudiant - au foyer - retraité).

Au total, 191 enquêteurs formés sont intervenus cette année, sur 282 points de chute. Des règles de dispersion strictes sont imposées : une seule personne interrogée par foyer, une seule personne interrogée par entrée d'immeuble, un pas de 4 est observé pour les maisons individuelles, etc.

La collecte a eu lieu entre le 13 décembre 2014 et le 29 janvier 2015. Cette période (6 semaines) inclut les deux semaines de vacances scolaires d'hiver, et couvre également la période des attentats qui ont secoué le pays début janvier.

Les contrôles de qualité sont principalement de deux ordres. Les contrôles postaux ont pour but de vérifier la réalité de l'interview, sa durée, son lieu de réalisation et l'exactitude de certaines réponses. 400 lettres de contrôle ont été envoyées à des interviewés de l'ensemble des enquêteurs, avec un taux de retour de 52%. Les contrôles téléphoniques ont porté sur 396 questionnaires (19,8% de l'ensemble). Contrairement aux contrôles postaux, pour lesquels les interviewés ont été désignés de manière aléatoire, les contrôles téléphoniques ont été réalisés en fonction d'indicateurs de qualité (durée trop courte, informations manquantes, trop de « ne sait pas » ou de non réponse, imprécisions sur la profession notamment, etc.).

L'enquête en ligne

L'enquête en ligne a été réalisée auprès d'un échantillon représentatif de la population âgée de 18 ans et plus de 2 045 personnes. Le plan de sondage est strictement identique à celui imposé à l'enquête en face-à-face. L'échantillon devait respecter les mêmes critères de quotas (ZEAT, taille d'unité urbaine, sexe, âge, PCS) que l'échantillon interrogé en face-à-face.

C'est le Newpanel d'Opinionway qui a été utilisé. C'est un access panel d'environ 100.000 panelistes recrutés par différents modes (par téléphone, via un site dédié mais aussi sur fichiers d'adresses). Il a récemment été enrichi sur deux cibles : les seniors et les ouvriers. Le profil de l'ensemble de ces panelistes, à l'exception d'une sur-représentation en cadres et membres des professions libérales, est très proche de celui de la population totale (ce que l'on retrouve dans les résultats obtenus auprès de 2.045 panélistes ayant répondu à l'enquête). En moyenne, ces panelistes sont sollicités 2 à 3 fois par mois pour participer à des études variées. 304 personnes ont abandonné le questionnaire en cours de passation, soit un taux de 12,9%, supérieur à la moyenne habituellement mesurée par Opinionway (5 à 10%), en raison de la durée du questionnaire, inhabituelle pour une enquête online (51 minutes, plus courte que la durée moyenne en face-à-face). Ce sont plus souvent des femmes (58% vs 52% dans les répondants hommes), de jeunes (26% ont moins de 30 ans vs 17%) et de personnes en emploi (61% contre 53%), un peu plus souvent issues des catégories sociales supérieures (29% vs 25%). Les abandonnistes résident un peu plus souvent dans des agglomérations de moins de 20.000 habitants (47% vs 41%).

2 138 personnes ont, au total, répondu à l'enquête mais 93 interviews ont été annulées, en raison d'une durée de passation trop faible. Le taux de retour final (personnes ayant répondu complètement vs personnes contactées) est de 13,8%.

II. La comparaison des deux échantillons

Des écarts sur certaines variables descriptives hors quota

Le diplôme : Les enquêtés ont beau présenter les mêmes profils sociodémographiques quant au sexe, à l'âge, à l'inscription géographique ou à la profession et catégorie socioprofessionnelle, des écarts apparaissent sur le niveau de diplôme. Si la proportion de bacheliers est très proche dans les deux échantillons, la différence se fait dans les niveaux inférieurs, où les enquêtés en face-à-face sont systématiquement surreprésentés, et les niveaux supérieurs sont surreprésentés dans l'échantillon online. Le redressement n'y change quasiment rien.

La taille du foyer : Les panélistes vivent moins souvent seuls (20% vs 30%) et, plus souvent dans un ménage de deux personnes (41% vs 32%). Leur structure est, de ce fait, plus proche de celle de la population totale.

Le statut d'occupation du logement : Les écarts sont également sensibles s'agissant de statut d'occupation du logement : on recense, en effet, 48% de propriétaires dans le panel en ligne (+ 12 points par rapport à l'enquête en face-à-face) et, également, un peu plus d'accédants (16% vs 12%).

Les revenus : Le dernier aspect abordé dans cette partie concerne les **revenus**, qui sont mesurés dans l'enquête de façon détaillée, par le biais de 6 questions balayant les types de ressources possibles et la personne qui, au sein du foyer, les perçoit. Dans l'enquête en ligne, on observe un niveau de revenus supérieur à ce qui est déclaré dans l'enquête en face-à-face : le niveau de vie médian est **supérieur de 20%** à ce qu'il est dans la population interrogée en face-à-face (1800 € vs 1500 €). Les panélistes en ligne sont ainsi 30% à avoir un niveau de vie supérieur à 2700€ net par mois, contre 18% dans l'enquête en face-à-face.

De nombreux indicateurs sur lesquels les deux enquêtes s'accordent

Sur de nombreuses questions, les résultats issus des deux enquêtes sont tout-à-fait comparables, et ce sur de multiples thèmes : *le rapport à l'emploi, le logement et son équipement, le cadre de vie, l'état de santé perçue, l'union et les possibilités d'adoption des couples de même sexe, l'opinion sur les causes individuelles de pauvreté et le chômage, la mondialisation et les échanges commerciaux, l'opinion sur l'augmentation probable du chômage et l'évolution de leur niveau de vie personnel et de celui des français, et enfin l'opinion sur les grandes marques commerciales.*

Des écarts sur certaines variables d'intérêt

Des écarts sur les nouvelles technologies, aussi bien sur les équipements que les usages : Les écarts sont, ici, attendus : les personnes recrutées dans le panel en ligne ont des taux d'équipement plus élevés, que ce soit en ordinateur (90% vs 81%), en connexion internet à domicile (97% vs 82%). Les écarts sont nets s'agissant d'équipements nomades que ce soit en tablette (10 points d'écart) ou en smartphone (16 points d'écart). Mais rappelons-le, les écarts d'équipement en téléviseur couleur et en téléphone mobile sont négligeables.

Ces différences se retrouvent quand on examine les usages numériques : les enquêtés du panel sont, par exemple, plus souvent membres d'un réseau social sur internet (+ 9 points) et, plus fréquemment, ils naviguent sur internet à partir d'un téléphone mobile (+ 7 points). Les différences pour les achats en ligne sont très nettes : la proportion de personnes ayant fait des achats en ligne s'élève à 87% dans le panel online (soit 34 points de plus qu'en face-à-face). Et, très souvent, pour les panélistes, internet sert à réaliser des économies, en permettant de procéder à des comparaisons entre produits et services en amont de l'achat (79%). Cette attitude est revendiquée par moins de la moitié de l'échantillon interrogé en face-à-face (46%, soit 33 points d'écart).

Des panélistes un peu moins inquiets que les personnes interrogées en face-à-face : Les panélistes témoignent d'une plus grande sérénité. Interrogés sur les inquiétudes que génèrent un certain nombre de risques aussi bien individuels (chômage, maladie grave, agression dans la rue, risques alimentaires) que collectifs (accident de centrale nucléaire, guerre), ils se montrent systématiquement un peu moins inquiets que les personnes interrogées en face-à-face. Limités pour les risques alimentaires (3 points), les écarts sont de 5 à 7 points pour le risque d'agression dans la rue, de maladies graves et d'accident de la route. Et ils sont de 10 points environ pour le risque de guerre et d'accident de centrale nucléaire, pour culminer à 20 points pour **le chômage** (57% d'inquiétude parmi les panélistes vs 78% pour les personnes interrogées en face-à-face), probablement en liaison avec la sur représentation de diplômés dans le panel.

Des panélistes aux réponses moins « extrêmes » concernant les risques perçus : Un des enseignements de la comparaison des deux enquêtes réside dans le fait que les modalités les plus extrêmes sont moins souvent utilisées par les enquêtés online. Cela se vérifie sur plusieurs types de variables, comme les notes de 1 à 7 ou les questions qui font référence à une échelle d'inquiétude en 4 niveaux : beaucoup, assez, un peu et pas du tout. Il s'avère que, bien souvent, les différences sont beaucoup plus marquées avec l'échelle détaillée. Pour certains risques (maladie grave, agression dans la rue et accident de la route), les écarts sont sensibles avec l'échelle

détaillée mais ils disparaissent lorsqu'on regroupe les réponses. Enfin, pour d'autres risques, les écarts subsistent même quand on regroupe les réponses. C'est le cas pour le chômage, le risque d'accident nucléaire ou le risque de guerre.

Des panelistes moins compatissants, plus racistes et aussi moins heureux : Les panélistes se disent moins souvent « pas du tout raciste » que les personnes interrogées en face-à-face (43% vs 58%, soit 15 points de moins). Dans le même esprit, ils disent plus souvent qu'ils souhaitent le départ d'un grand nombre d'immigrés qui vivent en France (17 points d'écart). Leur taux de réponse aux questions partisans est également plus élevé : 3% ont refusé de se situer sur une échelle politique contre 16% des enquêtés en face-à-face.

Aussi, quand il s'agit de dire si on est heureux ou pas dans la vie qu'on mène, les panélistes avouent plus facilement ne l'être qu'occasionnellement (33% vs 16%), alors que, face à un enquêteur, la tentation est plus grande d'assurer l'être « très souvent » (31% vs 14%).

Les panélistes se différencient également par le regard qu'ils portent sur la pauvreté, ses causes et les effets des politiques sociales. Sur certaines de ces questions, ils témoignent d'une position un peu moins compatissante vis-à-vis des personnes en bas de l'échelle sociale. Ainsi, ils pensent plus souvent qu'il est parfois plus avantageux de percevoir des minima sociaux que de travailler (77% vs 67%) et enfin qu'il y a trop des prestations familiales qui ne sont versées qu'aux plus démunis (58% vs 49%). Rappelons pourtant que sur des indicateurs proches de ceux-ci (sur les raisons de la pauvreté et le chômage), les deux échantillons ne faisaient pas apparaître de différence significative.

Synthèse des écarts pour les questions généralistes

Au final, le questionnaire Conditions de vie et Aspirations, au début 2015, comptabilise plus de 300 variables. Nous avons sommé tous les écarts en valeur absolue sur toutes les modalités présentes, et en avons calculé une moyenne pour chacun des onze thèmes du questionnaire (comprenant de 9 à 17 questions). Lorsqu'on mesure les écarts sur l'ensemble du corpus généraliste de questions entre les deux enquêtes menées en parallèle, on aboutit à un écart (en valeur absolue) de 3,9 points.

Cet écart est dépendant des thèmes abordés : il varie ainsi de moins de 2 points pour la situation d'emploi à presque 8 points pour les questions relatives aux questions sur les technologies de l'information (Tableau 1). Mais, il reste, somme toute relativement limité.

Tableau 1 – Indicateur résumé des écarts entre les deux modes de collecte, par bloc de questions

	Ecart moyen
Situation d'emploi	1,6
Santé	2,8
Moral économique	2,9
Opinions sur la famille	3,4
Sociodémographique	3,4
Logement	3,6
Valeurs	4,2
Opinions sur la société	4,7
Sociabilité	5,0
Opinions sur la solidarité	5,2
TIC	7,7
	3,9

Source : CREDOC, enquêtes Conditions de vie et Aspirations, début 2015.

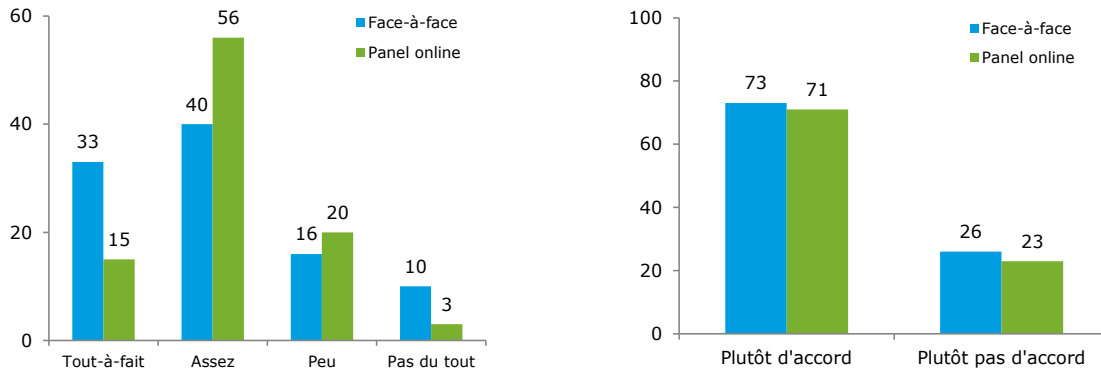
Exemple de lecture : en moyenne, pour chaque modalité des questions relatives à la sociabilité de l'enquêté, l'écart entre les réponses des enquêtés en face-à-face et les enquêtés en ligne est de 5 points (en valeur absolue).

Des écarts importants pour la désirabilité sociale et la présentation de soi

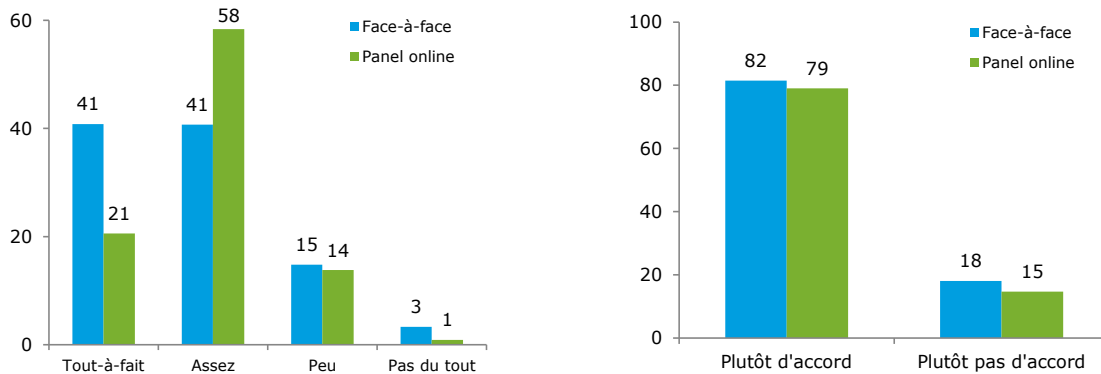
Les écarts entre panélistes et répondants en face-à-face sont importants sur les variables de désirabilité sociale ou de satisficing comme le montrent les figures suivantes, obtenues après calage (hypothèse 0 dans ce qui suit). Cela souligne le potentiel intérêt qu'il y a de les prendre en compte pour neutraliser une partie des effets modes de collecte.

Figure 1: Distribution des variables de désirabilité sociale

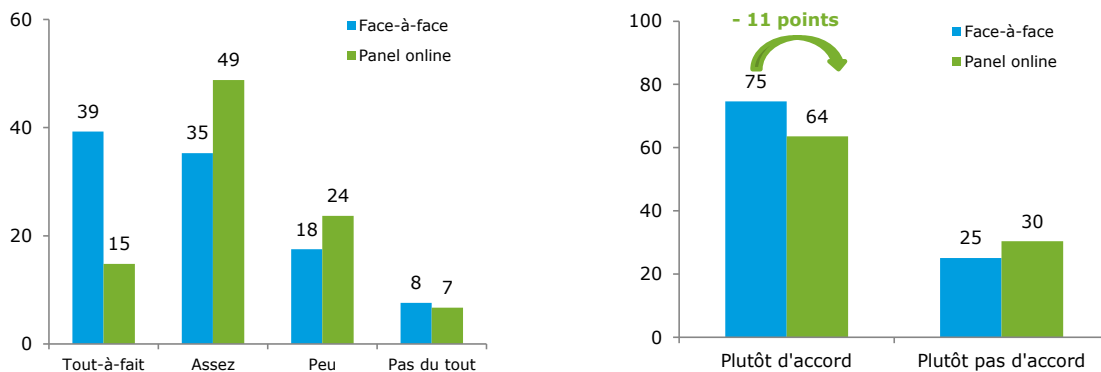
Dans la vie, vous faites toujours attention à ce que vous dites et ce que vous faites pour ne pas froisser les gens



Dans la vie, quoi que vous fassiez, vous êtes toujours très exigeant avec vous-même



Dans la vie, vous vous comportez toujours de la même façon lorsque vous êtes seul ou avec les autres



Source : CREDOC, enquêtes Conditions de vie et Aspirations, début 2015. Des écarts qu'il est difficile de corriger

III. Au-delà du calage simple, les tentatives de repondérations

Est-il possible de réduire cet écart moyen grâce à un redressement qui permettrait de mieux ajuster les répondants en ligne à la population interrogée en face-à-face ? Pour le savoir, nous avons testé plusieurs modèles de redressement et mesuré, à nouveau, l'écart moyen entre les deux modes de collecte pour voir si les écarts s'amenuisaient sur chacune des dimensions abordées dans l'enquête.

Le redressement « classique » de l'enquête est le résultat d'une post-stratification par calage sur marges (12) sur les principales caractéristiques sociodémographiques (âge x sexe, type de logement, PCS et taille d'agglomération). Il s'agit donc de l'hypothèse 0. Nous avons d'abord tenté d'améliorer cette technique en prenant en compte, dans le calage sur marges, des variables supplémentaires dont nous avons pu tester la pertinence. Successivement, ont donc été ajoutées aux quatre variables initiales les variables suivantes :

- Hypothèse 1 : L'utilisation d'internet pour comparer les prix des biens et services avant l'achat
- Hypothèse 2 : Le niveau de diplôme en 4 postes
- Hypothèse 3 : Les trois questions relatives à la désirabilité sociale (dont le détail est donné page XXX)
- Hypothèse 4 : La typologie des niveaux de vie, avec les découpages fournis à partir de l'enquête en face-à-face
- Hypothèse 5 : L'ensemble de toutes ces variables supplémentaires

Ensuite, d'autres tentatives ont consisté à estimer un score de propension(13). Ce score de propension détermine la probabilité, pour un individu, de participer au panel online. Il est estimé à partir des deux échantillons (individus interrogés en face-à-face + individus interrogés dans le panel), grâce à une régression logistique : ce score correspond à la probabilité, pour un individu, d'avoir participé à l'enquête online plutôt qu'à l'enquête en face-à-face. Pour construire ce score, on choisit des variables dites auxiliaires qui peuvent expliquer qu'une personne est plus disposée à répondre à l'enquête en ligne plutôt qu'en face-à-face. Les 11 variables suivantes ont été utilisées :

- la situation d'emploi,
- le diplôme,
- la typologie des niveaux de vie avec les découpages fournis à partir de l'enquête en face-à-face,
- le recours à internet pour comparer les offres avant achat
- les trois indicateurs de désirabilité sociale,
- la participation à un réseau social en ligne, l'achat en ligne, la connexion à internet sur téléphone mobile
- l'opinion sur le versement des prestations familiales aux plus démunis.

Nous avons exclu les variables de quota, équivalentes dans les deux échantillons. Nous avons choisi une option de sélection automatique des variables significatives du modèle (méthode Stepwise). Le modèle aboutit à un résultat 85,3% d'affectations correctes.

Une fois ce score de propension estimé, il existe plusieurs méthodes dans la littérature pour l'utiliser(14). Nous avons choisi d'opter pour l'inverse du score de propension. Cette méthode est assez intuitive : plus un individu a une probabilité faible d'avoir été interrogé sur le web plutôt qu'en face-à-face, plus on donnera un poids élevé à ses réponses afin de se rapprocher des réponses du face-à-face. Trois tests ont ensuite été menés :

- Hypothèse 6 : le poids de chaque individu est l'inverse du score de propension
- Hypothèse 7 : calage sur marge avec les 4 variables « classiques », avec comme poids initial du calage sur marges l'inverse du score de propension (et pas la valeur 1)
- Hypothèse 8 : calage sur marge en ajoutant la variable diplôme aux 4 variables « classiques », avec comme poids initial du calage sur marges l'inverse du score de propension (et pas la valeur 1)

Le **Tableau 2** récapitule l'ensemble des 8 hypothèses de pondération testées et les caractéristiques principales des poids obtenus. En effet, dans tout redressement, il existe un arbitrage entre la qualité du redressement (ici l'écart entre les résultats des deux modes de collecte) et la dispersion des poids qui est liée à une augmentation de la variance. Ici, l'hétérogénéité des poids augmente avec le nombre de marges sur lesquelles s'effectue le

calage. En particulier, les ratios de dispersion (poids maximal / poids minimal) mesurés pour les hypothèses 3 et 5 semblent quelque peu réhivitoires.

- L'hypothèse 0 aboutit à un écart moyen de 3,9 points, avec un écart par bloc allant de 1,6 points pour la santé à 7,7 points pour les questions relatives aux TIC.
- L'hypothèse 1 amplifie la distribution des poids (le ratio poids max / poids min est multiplié par 2) ; elle améliore la situation sur 6 des 11 blocs mais la dégrade sur 4 autres.
- L'hypothèse 2 accentue encore la distribution des poids (le ratio poids max / poids min est à nouveau doublé) ; elle améliore la situation sur 7 des 11 blocs mais la dégrade sur 3 autres.
- L'hypothèse 3 doit être rejetée au motif que les poids obtenus sont beaucoup trop dispersés.
- L'hypothèse 4 améliore la situation (sur 5 blocs) autant qu'elle l'empire (sur 5 autres).
- L'hypothèse 5 doit également être rejetée au motif que les poids obtenus sont beaucoup trop dispersés.
- L'hypothèse 6 améliore la situation quant aux thèmes sur lesquels les écarts étaient les plus importants (sociabilité, opinions sur la société et TIC) mais elle la dégrade sur 5 autres thèmes, dont la situation d'emploi et les descripteurs sociodémographiques.
- L'hypothèse 7 réduit les écarts sur 7 des 11 blocs, mais elle génère des poids très dispersés (ratio de 160 entre le poids minimal et le poids maximal).
- L'hypothèse 8, quant à elle, réduit les écarts sur 8 des 11 blocs étudiés, mais au prix d'une plus grande dispersion encore (ratio de 303 entre le poids minimal et le poids maximal).

Quelle conclusion en tirer ?

Sur les neuf essais, deux au moins sont à proscrire car les poids obtenus sont trop hétérogènes ; ce sont pourtant les deux qui aboutissent à la meilleure réduction des écarts (3,5 points en moyenne). Pour les autres, les améliorations sont perceptibles mais pas systématiques : **aucun scénario ne permet, en effet, de diminuer les écarts sur toutes les thématiques de questions simultanément**. Au final l'écart moyen est ramené, au mieux, à 3,6 points (contre 3,9 points avec un redressement « classique »).

I. Une tentative d'appariement pour éliminer le biais de sélection et de désirabilité sociale

Deux méthodes principales s'affrontent dans le redressement des enquêtes multimodes non probabilistes (15) : le calage sur marges (12, 16) et l'usage de scores de propension (17, 18). Ces deux méthodes peuvent éventuellement être combinées, comme cela a été fait précédemment. Aucune des deux ne permet de corriger tous les biais (19), mais théoriquement, la conjonction des deux permet de tenir compte de davantage de caractéristiques, ce qui doit la rendre plus efficace, même si, comme nous venons de le voir, le résultat n'est pas tout à fait concluant ici. Cette méthodologie se heurte toutefois à une limite : les individus que l'on pondère peuvent être trop différents d'un mode de collecte à l'autre, ce qui peut, outre une inflation de la variance, entraîner la comparaison d'individus non comparables car très dissemblables. Le même problème se pose si l'on utilise le score de propension non pas pour redresser l'échantillon mais pour mesurer une différence entre échantillons ajustée de la propension à répondre sur chaque mode de collecte (18, 20). Cela est en partie évitable : pour cela, on restreint l'analyse aux individus présentant des caractéristiques semblables.

Pour aller plus loin, nous avons donc opéré un appariement un à un des individus des deux échantillons présentant un même score de propension (la probabilité qu'un individu ait répondu sur un mode donné, par exemple Internet). Cet appariement un à un équilibre en principe parfaitement, asymptotiquement, les caractéristiques introduites dans la modélisation du score dans les deux groupes appariés (17, 21). Les catégories de variables introduites dans le score de propension sont les suivantes :

Tableau 2 – Modalités des essais de redressement et caractéristiques des poids obtenus

	Poids initial	Calage sur les marges suivantes :		Caractéristiques des poids :				Ecart moyen après redressement
				Min	Max	Max/Min	Médiane	
Redressement classique sur l'enquête en face-à-face	1	Sexe x âge/PCS Type de logement Taille d'agglom.		0.46	4.20	9.13	0.95	3.9
Hypothèse 0	1	Sexe x âge/PCS Type de logement Taille d'agglom.		0.40	7.79	19.48	0.86	3.9
Hypothèse 1	1	Sexe x âge/PCS Type de logement Taille d'agglom.	+ A utilisé internet pour comparer les prix avant l'achat	0.23	10.88	47.30	0.58	3.8
Hypothèse 2	1	Sexe x âge/PCS Type de logement Taille d'agglom.	+ Diplôme	0.15	14.17	94.47	0.70	3.8
Hypothèse 3	1	Sexe x âge/PCS Type de logement Taille d'agglom.	+ 3 questions sur la désirabilité sociale	0.01	14.68	1468.00	0.61	3.5
Hypothèse 4	1	Sexe x âge/PCS Type de logement Taille d'agglom.	+ Typologie des niveaux de vie	0.21	8.86	42.19	0.84	3.9
Hypothèse 5	1	Sexe x âge/PCS Type de logement Taille d'agglom.	+ Diplôme + A utilisé internet pour comparer les prix + 3 questions désirabilité sociale + Typologie niveaux de vie	0.002	31.96	15980.00	0.38	3.5
Hypothèse 6	1/ score de propension	-	-	20.45	1786.72	87.37	27.19	3.7
Hypothèse 7	1/ score de propension	Sexe x âge/PCS Type de logement Taille d'agglom.		0.17	27.20	160.00	0.65	3.6
Hypothèse 8	1/ score de propension	Sexe x âge/PCS Type de logement Taille d'agglom.	+ Diplôme	0.12	36.40	303.33	0.57	3.7

Source : CREDOC, enquêtes « Conditions de vie et aspirations », début 2015.

Sociodémographiques : Sexe, âge (5 catégories), diplôme (7 catégories), couple (oui/non), type de logement (3 catégories), taille d'agglomération (5 catégories), tranche de revenu (7 catégories), situation professionnelle (3 catégories), a des enfants (oui/non).

Webographiques (i.e. les variables caractérisant l'équipement et l'usage de matériel connecté) : Possession d'un micro-ordinateur (3 catégories), possession d'une tablette (oui/non), domicile connecté à Internet (oui/non), possession d'un smartphone connecté (2 catégories), quintile d'heures passées sur Internet (5 catégories).

Enquêtographiques (i.e. les variables caractérisant la participation à des enquêtes passées) : A participé à des enquêtes sur Internet (oui/non) ; a participé à des enquêtes au téléphone (3 catégories : 0, 1, 2+) ; a participé à des enquêtes en face-à-face (3 catégories : 0, 1, 2+). La participation à des enquêtes sur Internet (oui/non) a été testée mais non prise en compte directement car les réponses positives étaient quasi-exclusivement le fait des panélistes. En revanche, une ACP sur ces trois variables permet de retenir les deux premières composantes, expliquant 76.8% de la variance. Nous utiliserons ces deux composantes dans le score de propension.

Ipsographiques (i.e. les variables relatives à la présentation de soi), dont les intitulés exacts sont : Vous êtes superstitieux ; Vous êtes prudent ; Dans la vie, vous faites toujours attention à ce que vous dites et à ce que vous faites pour ne pas froisser les gens ; Dans la vie, vous vous comportez de la même manière lorsque vous êtes seul ou avec les autres ; Dans la vie, quoi que vous fassiez, vous êtes toujours très exigeant avec vous-même. Les modalités de réponse sont identiques pour toutes les questions : tout à fait d'accord, plutôt d'accord, plutôt pas d'accord, pas du tout d'accord. La modalité initiale NSP et les non-réponses, en nombres très faibles, ont été recodées par la méthode d'imputation hotdeck.

Il est à noter que seules les trois dernières questions proviennent d'une synthèse, très libre (compte tenu du manque de place dans le questionnaire) des échelles de désirabilité sociale de satisficing citées précédemment. Mais toutes sont relatives au comportement et à la présentation de soi, ce qui est susceptible d'être lié au comportement de réponse aux variables d'intérêt.

Variables d'intérêt : Les variables d'intérêt étudiées ici sont au nombre de 6 : tabagisme quotidien, consommation régulière d'alcool, expérimentation de cannabis, jugement négatif sur son cadre de vie, sentiment d'être en mauvaise santé, conduite automobile. Ces variables traduisent des comportements plus ou moins dévalorisés socialement (expérimentation de cannabis, tabagisme quotidien, consommation régulière d'alcool), neutres (conduite automobile), ou bien relatifs à la perception d'une mauvaise qualité de vie (santé ou cadre de vie dégradés). À ce titre, elles exemplifient une palette de comportements plus ou moins socialement valorisés et stigmatisés.

Au total, six scores de propension ont été estimés par régression logistique, définis par les variables suivantes :

- sociodémographique +/- ipsographiques (psd, psd_i) ;
- sociodémographique + webographique +/- ipsographiques (psdw, psdw_i) ;
- sociodémographique + webographique + enquêtographique +/- ipsographiques (psdwe, psdwe_i).

Le nombre de panélistes appariés ainsi que l'aire sous la courbe ROC du modèle (22) est fourni tableau 3. Plus l'ajustement comporte de variables, plus le nombre d'individus appariés est réduit.

Tableau 3 - Résumé des modélisations de la propension (probabilité associée à la réponse sur le panel Internet) pour différents ajustements

	Sociodémographique (psd)	sociodémographique + webographique (psdw)	sociodémographique + webographique + enquêtographique (psdwe)
Sans ipsographiques	C=0.728 ; n=2714 / n=90*	C=0.793 ; n=2244 / n=30*	C=0.926 ; n=980 / n=16*
Avec ipsographiques	C=0.811 ; n=2136 / n=36*	C=0.851 ; n=1898 / n=94*	C=0.942 ; n=936 / n=42*

C=aire sous la courbe ROC

*=nombre d'individus appariés sur une décimale.

Tableau 4 - Odds-ratios bruts et intervalles de confiance à 95% pour l'association entre la réponse sur Internet et quelques variables d'intérêt pour différents modèles d'appariement

	Brut pondéré par le calage Hypothèse 0 n=4045			PSD n=2714			PSD_i n=2136			PSDWE n=980			PSDWE_I n=936		
	OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL
Cannabis vie	0,71	0,61	0,82	0,64	0,53	0,76	0,59	0,48	0,72	0,47	0,34	0,63	0,51	0,38	0,69
Tabac quotidien	0,43	0,37	0,50	0,45	0,37	0,54	0,51	0,41	0,62	0,53	0,39	0,71	0,51	0,37	0,69
Alcool régulier	1,06	0,91	1,24	0,96	0,79	1,16	0,89	0,72	1,10	0,99	0,71	1,38	0,99	0,71	1,38
état de santé dégradé	1,06	0,90	1,25	1,44	1,17	1,77	1,42	1,13	1,79	1,29	0,91	1,84	1,49	1,05	2,11
cadre de vie dégradé	0,99	0,83	1,19	1,22	0,98	1,52	1,38	1,08	1,76	1,55	1,08	2,21	1,39	0,96	2,01
Conduite automobile	2,12	1,80	2,49	1,54	1,25	1,90	1,33	1,06	1,67	1,09	0,78	1,52	1,12	0,79	1,60

Légende : PSD : ajustement sur les variables sociodémographiques ; PSD_i : ajout des variables ipsographiques dans le calcul du score de propension précédent ; psdwe : score de propension calculé sur les variables sociodémographiques, webographiques et enquêtographiques ; PSDWE_i : ajout des variables ipsographiques dans le calcul du score précédent. OR=odds-ratio bruts ; LCL=lower confidence limit ; UCL=upper confidence limit.

Le passage d'une comparaison sur les échantillons initiaux pondérés par le calage de l'hypothèse 0 aux comparaisons sur les modèles appariés ne révèle pas toujours de convergence des mesures : l'écart augmente pour l'état de santé par exemple. L'ajout des variables ipsographiques dans un modèle d'appariement accroît parfois les écarts (passage de PSD à PSD_i pour l'alcool et de PSDWE à PSDWE_i pour l'état de santé). Ces résultats corroborent les résultats bi-variés obtenus précédemment. Les résultats obtenus pour le modèle d'appariement ajusté PSDW sont similaires. La seule variable pour laquelle les appariements successifs apportent une convergence des prévalences dans les deux échantillons est la conduite automobile.

L'étude n'apporte donc pas de preuve que l'ajout de variables attitudinales diminue les écarts relativement à l'ajustement sur des variables plus objectives classiques.

II. L'extrapolabilité des résultats : une comparaison de plusieurs panels

L'étude montre que les prévalences relevées dans le panel et en face-à-face sont très proches pour tout un ensemble de variables mais au contraire très dissemblables pour une série d'indicateurs comportementaux et qu'il est vain de vouloir réduire les écarts par des techniques de redressement ou même d'appariement.

Les variables attitudinales et notamment relatives à la présentation de soi (ipsographiques et relatives à la désirabilité sociale et au satisficing) semblent peu utiles dans ces tentatives. Cela pourrait être dû au fait que les questions correspondantes sont également victimes de biais de désirabilité sociale et d'effet modes de collecte... toutefois, la question de l'extrapolabilité des résultats de la comparaison à d'autres panels se pose. Les caractéristiques du panel utilisé ici (conservatisme, préférence pour certaines affirmations marquées de valeurs liées au social et à la politique etc.) pourraient peut-être expliquer une partie de nos résultats. Une étude récente comparant une enquête probabiliste à plusieurs access panels a en effet conclu à la variabilité des écarts suivant le panel et les quotas utilisés (23). Pour tester ce point, nous avons opéré une comparaison brève trois panels : Toluna (n=1000), Panel On the Web (n=1000), Opinion Way (n=3050). Les panélistes y ont été recrutés suivant les mêmes quotas.

Panel on the web a constitué son échantillon à partir d'une base de 100 000 internautes, gérée par leur soin avec une actualisation annuelle des informations. Toluna a pour sa part constitué son échantillon dans sa propre base de 290 000 internautes en France. La collecte a eu lieu entre le 22 décembre 2015 et le 29 janvier 2016, soit une durée totale de 39 jours. OpinionWay a collecté ses données entre le 21 décembre 2015 et le 24 janvier 2016, soit une durée totale de 35 jours.

Le questionnaire administré intégrait:

- Des **variables sociodémographiques simples** : sexe, âge, profession, niveau d'éducation de l'individu, région et taille d'agglomération dans laquelle réside l'interviewé ainsi que le statut d'occupation et type du logement, l'anticipation des conditions de vie personnelles, la situation actuelle de l'interviewé (actif, étudiant, chômeur, ...), le nombre d'individus vivant dans le logement et enfin les revenus de l'individu.
- Des **variables d'intérêt** : la sensibilité à l'environnement, la perception de de l'état de santé personnel, l'obligation de restrictions dans le budget, le départ en vacances au cours des 12 derniers mois, l'utilisation d'internet pour réaliser des économies, l'évaluation de certaines prestations familiales (RMI, RSA) et leurs conséquences, l'attrance pour les nouvelles technologies, l'opinion sur le fonctionnement de la justice, la nécessité d'une transformation de la société française, si oui, les types de réformes à implémenter, la confiance donnée plutôt à l'Etat ou aux entreprises pour régler des difficultés économiques, le sentiment de bonheur personnel, la classe politique et enfin l'opinion sur l'immigration dans les prochaines années.

Notons que, dans le cas d'OpinionWay, ces questions s'inséraient dans le questionnaire total de l'enquête 2016 sur les Conditions de vie et Aspirations des Français (d'une durée moyenne de 60 minutes en ligne avec un taux de réponse de 27%).

La durée moyenne pour le remplissage du questionnaire mesurée par Toluna était de 6 minutes et 25 secondes. Il ne dépassait pas 7 mn.

Plan de sondage

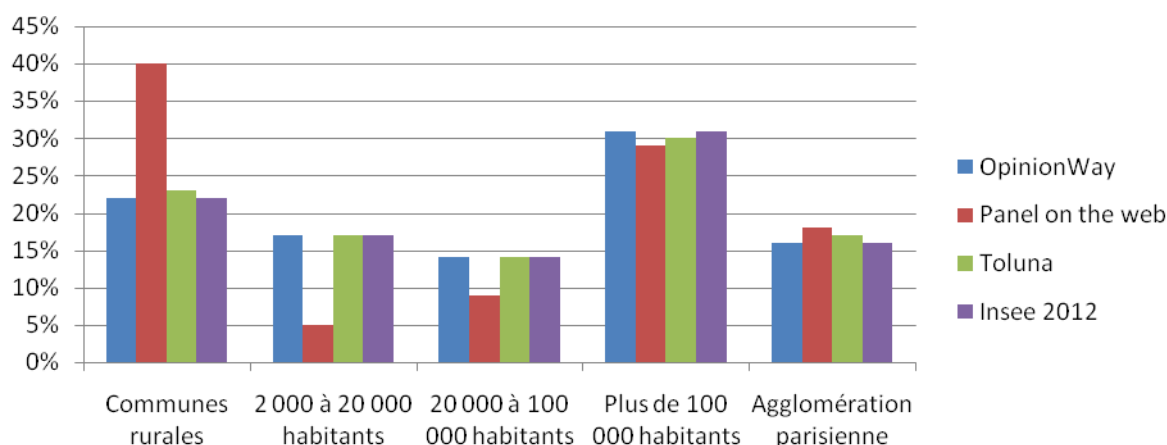
Les tailles d'échantillon demandées étaient respectivement de 3 050 individus pour OpinionWay et 1 000 individus pour Panel on the web et Toluna. Les échantillons sont représentatifs des individus habitant en France métropolitaine, Corse et DOM, âgés de 15 ans et plus, sélectionnés selon la méthode des quotas :

- Sexe
- Age en six (15-19 ans / 20-29 / 30-49 / 50-59 / 60-69 ans/ 70 ans et plus)
- Profession et catégorie sociale en huit catégories (travailleur indépendant, cadre et profession intellectuelle supérieure, profession intermédiaire, employé, ouvrier, personne au foyer, retraité et autre inactif)
- Taille d'unité urbaine en cinq (commune rurale, de 2 000 à 20 000 habitants, de 20 000 à 100 000 habitants, plus de 100 000 habitants et agglomération parisienne)
- Répartition par nouvelles grandes régions avec Corse et DOM
- Type de logement (individuel ou collectif).

De plus, il a été demandé à ce que le niveau de diplôme des répondants ne soit pas trop éloigné de la distribution en population générale (quota dit « soft »). Les quotas sur le sexe, l'âge, les professions et catégories sociales ainsi que sur le type de logement ont été respectés par les 3 sociétés de panel. Concernant le quota sur la répartition par nouvelles grandes régions (en 14 avec la Corse et les DOM), il a été globalement respecté par tous les prestataires. Néanmoins, seul OpinionWay a sondé suffisamment en Corse et dans les DOM, Toluna n'ayant aucun effectif dans ces deux régions.

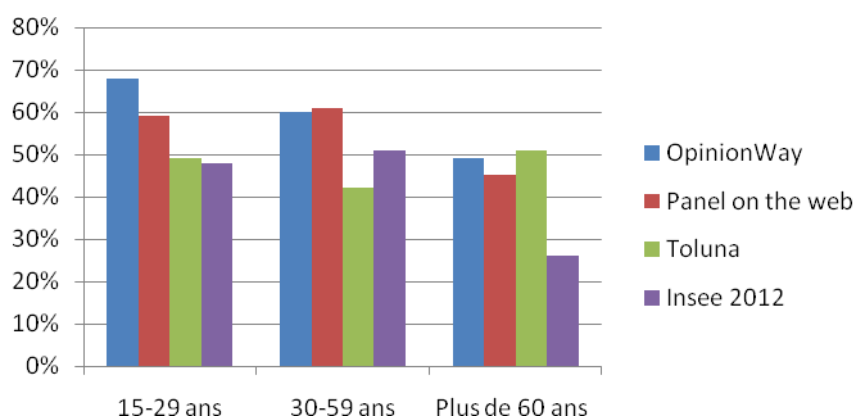
Enfin, le quota portant sur la répartition des populations par unité d'agglomération a bien été respecté par OpinionWay et Toluna mais pas du tout par Panel on the web. Ce dernier, malgré les quotas fixés, a sur représenté les habitants des « Communes rurales » et sous représenté les catégories « 2 000 à 20 000 » et « 20 000 à 100 000 ». Une distorsion qui n'a pas pu être corrigée par redressement. Nous avons étudié le niveau d'étude simplifié (en-dessous ou bien au-dessus du baccalauréat) par tranches d'âges : 15 à 29 ans, 30 à 59 ans et plus de 60 ans. Quelles que soient les sociétés de panel, on note une sur représentation marquée des diplômés, particulièrement marquée chez les 60 ans et plus.

Figure 2 - Répartition population par unité d'agglomération :



Source : CREDOC, expérimentation panels.

Figure 3 – Dernier diplôme obtenu supérieur ou égal au baccalauréat :



Source : CREDOC, expérimentation panels.

Redressement

Les 3 panels ont fait l'objet d'un redressement portant sur les variables suivantes :

- Sexe
- Taille d'unité urbaine en 5
- Profession et catégorie sociale en 8
- Nouvelles grandes régions en 14
- Type de logement (individuel, collectif)
- Age (moins de 20 ans, entre 20 et 29 ans, entre 30 et 49 ans, entre 50 et 59 ans, et plus de 59 ans) couplé avec le niveau d'éducation (inférieur au baccalauréat et supérieur ou égal au baccalauréat)

Le tableau 5 donne une première lecture des pondérations appliquées pour chaque société de panel et nous pouvons noter que **Panel on the web** se distingue avec un écart-type beaucoup plus élevé que les concurrents. La médiane ainsi que les extrêmes sont aussi particulièrement éloignés de ceux qui ont été appliqués aux échantillons d'OpinionWay et Toluna.

Tableau 5 – Description des coefficients de pondérations par société de panel

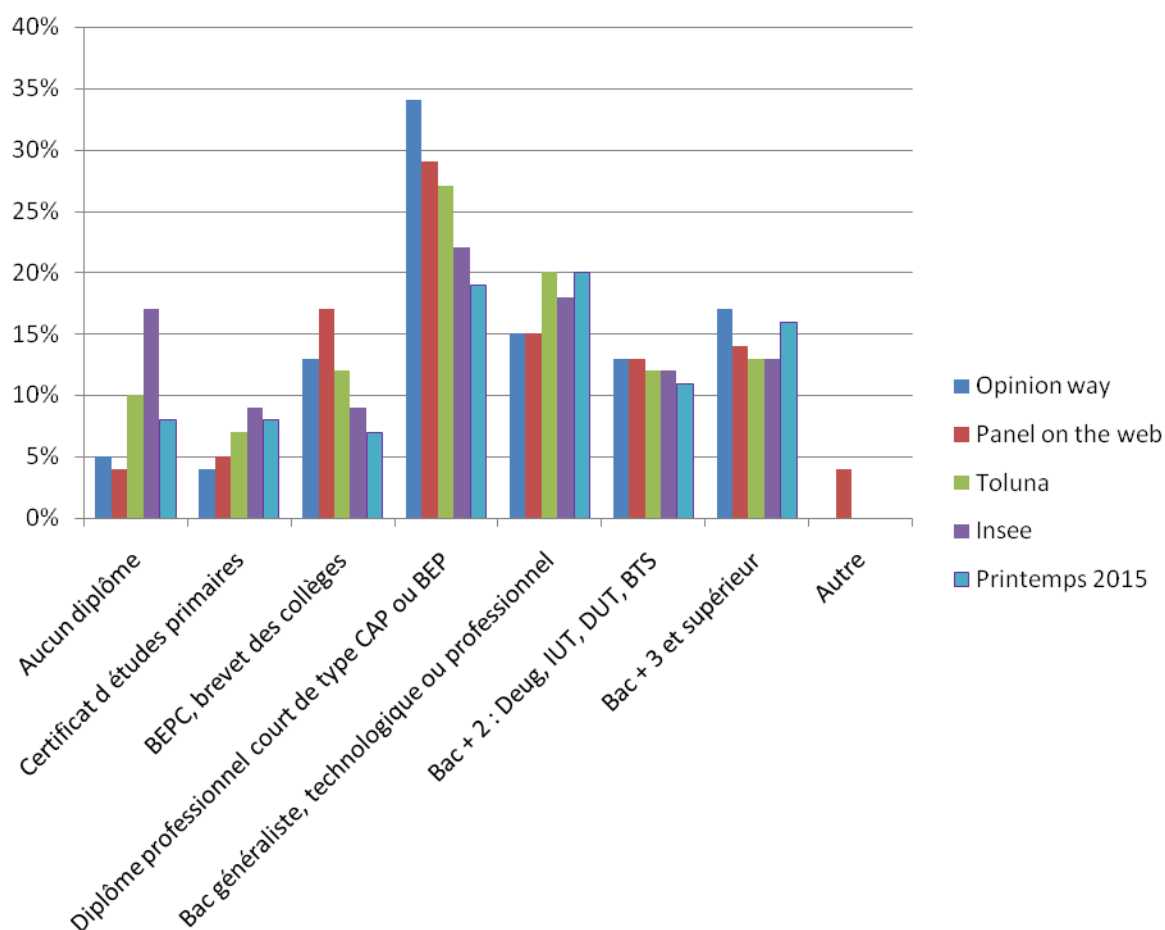
	N Obs	Moyenne	Médiane	Ecart-type	Minimum	Maximum
Opinion Way	3050	1,00	0,95	0,45	0,30	3,32
Panel on the Web	1000	1,00	0,79	0,74	0,17	6,54
Toluna	1000	1,00	0,95	0,39	0,26	2,71

Source : CREDOC, expérimentation panels.

I. Comparaison de la composition des trois panels

Le redressement ne corrige pas dans le détail la distribution des niveaux de diplôme détaillés : **les 3 panels sous-représentent les non diplômés et les CEP**. De plus, OpinionWay sur représente les « Bac+3 ». En sommant les écarts nets pondérés, nous trouvons que c'est **Toluna qui se rapproche le plus de la distribution moyenne de la population**.

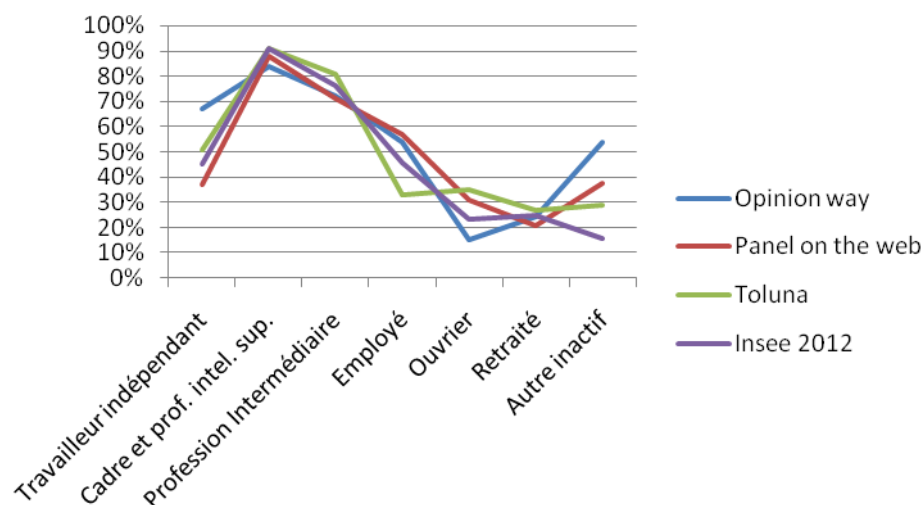
Figure 4 – Diplôme le plus élevé obtenu :



Source : CREDOC, expérimentation panels.

En étudiant le niveau de diplôme par profession et catégorie socioprofessionnelle, nous remarquons des écarts conséquents de la part des trois sociétés de panels par rapport à l'INSEE sur la catégorie « Autres inactifs ».

Figure 5 – Niveau de diplôme supérieur ou égal au baccalauréat :

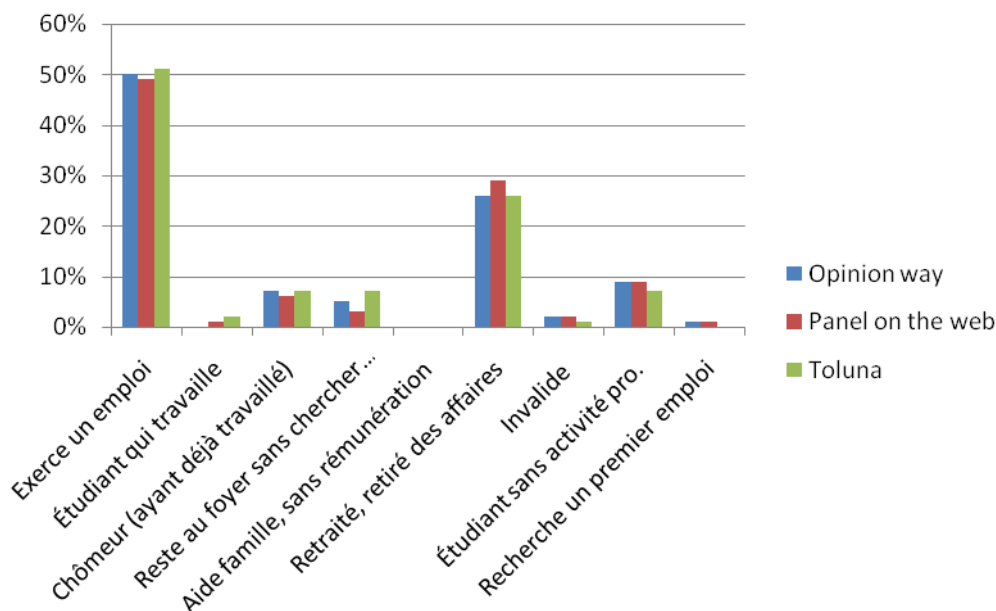


Source : CREDOC, expérimentation panels.

En étudiant le nombre de personnes vivant dans le même foyer, nous pouvons conclure que le panel **OpinionWay présente trop de personnes seules** : +6% par rapport aux données de l’Insee 2012 et aux autres panels.

Pour ce qui concerne la **situation actuelle** (actif, étudiant qui travail, chômeur, non demandeur d’emploi, aidant sa famille sans rémunération, retraité, invalide, étudiant sans activité professionnelle, demandeur d’un premier emploi), ou bien le **niveau de revenus** des interviewés, les trois panels sont **parfaitement comparables**.

Figure 6 – Situation actuelle des individus par panel :



Source : CREDOC, expérimentation panels.

L'analyse préliminaire suggère que les panels diffèrent pour certaines variables sociodémographiques hors quotas (taille de l'agglomération), les niveaux de diplôme, le statut d'occupation du logement, le nombre d'occupants du logement, la perception de l'état de santé, etc. Ces écarts ne sont toutefois pas très importants (quelques points tout au plus). Toutefois, il y a des écarts plus importants pour certaines expressions d'optimisme relatif à l'évolution de ses conditions de vie personnelles, la technologie dans les produits de consommation, l'expression de volonté de réforme de la société, les opinions politiques, la tolérance vis-à-vis des immigrés.

Dans des analyses multivariées ajustées sur l'âge, le diplôme, la perception de l'évolution de ses conditions de vie, les revenus, l'opinion politique et l'obligation de faire des restrictions budgétaires, il reste des écarts significatifs entre les panels pour quelques variables comme la perception de l'état de santé dégradé et même le fait d'utiliser le Web pour comparer les prix et faire des économies.

III. Conclusion

Ce travail montre l'existence de différences irréductibles entre les enquêtes effectuées sur un access panel et celles effectuées en face-à-face, même lorsque les deux reposent sur des quotas identiques. Elle suggère aussi que l'ajout de questions attitudinales liés à la présentation de soi, à la désirabilité sociale et au satisficing semble très peu contribuer à réduire les écarts observés dans des redressements aux variables sociodémographiques plus classiques. De même, le gain du recours préalable à des scores de propension avant calage semble discutable. Il montre aussi que les panels présentent des similarités sociodémographiques et d'expressions de valeurs notamment politiques et sociales, qui suggèrent une certaine extrapolabilité des résultats. Toutefois, des comparaisons plus approfondies sont en cours pour vérifier cette hypothèse.

Références

1. TOURANGEAU, R. & PLEWES, T. (2013) Nonresponse in social surveys: a research agenda *Committee on national statistics*, pp. 166 (Washington DC, The national academy press).
2. BIGOT, R. & CROUTTE, P. (2011) *La diffusion des technologies de l'information et de la communication dans la société française* (Paris, CREDOC).
3. GOMBAULT, V. (2013) L'internet de plus en plus prisé, l'internaute de plus en plus mobile, *Insee Première*, 1-4.
4. GOMBAULT, V. (2011) Deux ménages sur trois disposent d'internet chez eux, *Insee Première*, 1-4.
5. YEAGER, D. S., KROSNICK, J. A., CHANG, L. et al. (2011) Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples, *Public Opinion Quarterly*, vol. 75, pp. 709-747.
6. CHANG, L. & KROSNICK, J. (2010) Comparing oral interviewing with self-administered computerized questionnaires. An experiment., *Public Opinion Quarterly*, 74, 154-67.
7. DE LEEUW, E. D. (2008) Choosing the method of data collection, in: De Leeuw, E. D., Hox, J. J. & Dillman, D. A. (Eds.) *International handbook of survey methodology*, pp. 117-135 (New York, Lawrence Earlbaum Associates).
8. SCHONLAU, M., VAN SOEST, A. & KAPTEYN, A. (2007) Are "Webographic" or attitudinal questions useful for adjusting estimates from Web surveys using propensity scoring?, *Survey Research Methods*, 1, 155-163.
9. NENKOV, G. Y., MAURIN, M., WARD, A., SCHWARTZ, B. & HULLAND, J. (2008) A short form of the Maximization Scale: Factor structure, reliability and validity studies, *Judgment and Decision Making*, 3, 371-388.
10. KREUTER, F., PRESSER, S. & TOURANGEAU, R. (2009) Social desirability bias in CATI, IVR and Web surveys. The effect of mode and question sensitivity, *Public Opinion Quarterly*, 72, 847-865.
11. HOLBROOK, A. L., GREEN, M. C. & KROSNICK, J. A. (2003) Telephone versus face-to-face Interviewing of national Probability samples with long Questionnaires Comparisons of respondent satisficing And social desirability response bias, *Public Opinion Quarterly*, 79-125.
12. SAUTORY, O. (1993) *La macro Calmar. Redressement d'un échantillon par calage sur marges* (Paris, INSEE).
13. BIGOT, R., CROUTTE, P. & RECOURS, F. (2010) Enquêtes en ligne : peut-on extrapoler les comportements et les opinions des internautes à la population générale ? *Cahier de recherche* (Paris, Credoc).
14. VALLIANT, R. & DEVER, J. A. (2011) Estimating Propensity Adjustments for Volunteer Web Surveys, *Survey Research Methods*, 40, 105-137.
15. AAPOR (2013) Report of the AAPOR task force on non-probability sampling (AAPOR).
16. HAZIZA, D. & LESAGE, E. (2015) A discussion of weighting procedures for unit nonresponse, *Journal of Official Statistics*, to be published.
17. ROSENBAUM, P. R. & RUBIN, D. B. (1983) The central role of propensity scores in observational studies for causal effects, *Biometrika*, 70, 45-55.
18. ROSENBAUM, P. R. (1987) Model-based direct adjustment, *Journal of the American Statistical Association*, 82, 387-394.
19. BLOM, A. G. (2009) Nonresponse Bias Adjustments: What Can Process Data Contribute? *Institute for Social and Economic Research* (University of Essex).
20. RUBIN, D. B. (2001) Using propensity scores to help design observational studies: Application to the tobacco litigation, *Health Services & Outcomes Research Methodology*, 2, 169-188.
21. ROSENBAUM, P. R. & RUBIN, D. B. (1984) Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, *Journal of the American Statistical Association*, 79, 516-524.
22. REY, J. M., MORRIS-YATES, A. & STANISLAW, H. (1992) Measuring the accuracy of diagnostic tests using receiver operating characteristics (ROC) analysis, *International Journal of Methods in Psychiatric Research*, 1-11.
23. ERENS, B., BURKILL, S., COUPER, M. et al. (2014) Nonprobability Web Surveys to Measure Sexual Behaviors and Attitudes in the General Population: A Comparison With a Probability Sample Interview Survey, *Journal of medical internet research*, 16.