
COLLECTE MULTIMODE DE L'ENQUÊTE DEFIS SALARIES : A LA RECHERCHE D'UN EFFET DE MESURE

Edmond NOACK (*), Jean-Claude SIGOT (**)

(*) Cereq, Département formation et certification

(**) Cereq, Département formation et certification

noack@cereq.fr

Mots-clés: Multimode, effet de sélection, effet de mesure, effet de mode, appariement, score de propension

Résumé :

Cet article a pour objet l'analyse de biais liés au mode de collecte dans le cadre d'une enquête multimode.

Les données utilisées dans le cadre de cette analyse sont issues du dispositif DEFIS (Dispositif d'enquêtes sur les formations et les itinéraires des salariés) conçu géré par le Céreq pour le compte du CNEFP. Ce dispositif comporte un volet « entreprises » (les entreprises ont été interrogées au printemps 2015) couplé avec un volet « salariés » faisant l'objet d'interrogations annuelles s'étalant sur 5 ans (de 2015 à 2019).

C'est la 2^{ème} vague d'interrogation de ce panel salariés qui a été sollicitée dans le cadre de cette analyse. La collecte multimode qui a été mis en place comporte une phase de collecte par internet suivie en cas de non réponse par une phase de contacts téléphoniques pouvant déboucher sur une enquête internet ou téléphone. Ce mode de collecte fera l'objet d'une présentation succincte.

Afin de pouvoir mesurer d'éventuels biais liés à cette collecte multimode, une partie de l'échantillon, dit « échantillon de contrôle », s'est vu affecté aléatoirement à un mode de collecte spécifique (internet ou téléphone) afin de simuler une expérimentation randomisée.

Dans un premier temps l'analyse se portera uniquement sur les répondants en première intention de l'échantillon de contrôle. Nous nous sommes concentrés sur des variables d'opinion sensibles susceptibles d'être affectées par le mode de collecte. A l'aide de méthodes de régression, l'impact du mode de collecte sur ces réponses à l'enquête a été isolé. Les distributions de ce type de variables, conditionnellement à des variables sociodémographiques, ont été comparées entre les réponses de l'échantillon de contrôle ayant répondu par téléphone et celui par internet.

Puis dans un deuxième temps, en s'intéressant cette fois-ci à l'ensemble des répondants à l'enquête, nous avons utilisé une méthode de matching avec score de propension. Cette méthode permet d'annuler l'effet de sélection en créant pour chaque individu traité, répondant par internet (T=1 ou 0) selon que l'individu soit répondant par internet ou par téléphone) un contrefactuel qui est recherché et sélectionné parmi les non traités selon les variables qui explique la sélection (X).

$$Y = T * Y1 + (1 - T) * Y0$$

Abstract:

The purpose of this article is to analyze collection bias in a multi-mode survey.

The data used in this analysis are derived from the DEFIS system (Employee Training and Routing Survey System) managed by Céreq.

This was the second wave of questions from this employee panel that was asked as part of this analysis. The multimode collection that has been set up includes an Internet collection phase followed in case of non-response by a telephone contact phase that can lead to an Internet or telephone survey.

In order to measure possible biases related to this multimode collection, part of the sample, known as the "control sample", was randomly assigned to a specific collection mode (Internet or telephone) in order to simulate a randomized trial.

Initially, only first-line respondents will be analyzed to simulate randomization. We focused on sensitive opinion variables that could be affected by the collection method. Using regression methods, the impact of the collection mode on these survey responses was isolated. The distributions of this type of variable, conditional on socio-demographic variables, were compared between the responses of the control sample that responded by telephone and those by Internet.

Then in a second step, this time focusing on all respondents to the, we used a matching method with propensity score. This method makes it possible to cancel the selection effect by creating for each individual treated, responding by internet (T=1 or 0 depending on whether the individual is responding by internet or by telephone) a counterfactual which is searched and selected among the non-treated according to the variables which explains the selection (X).

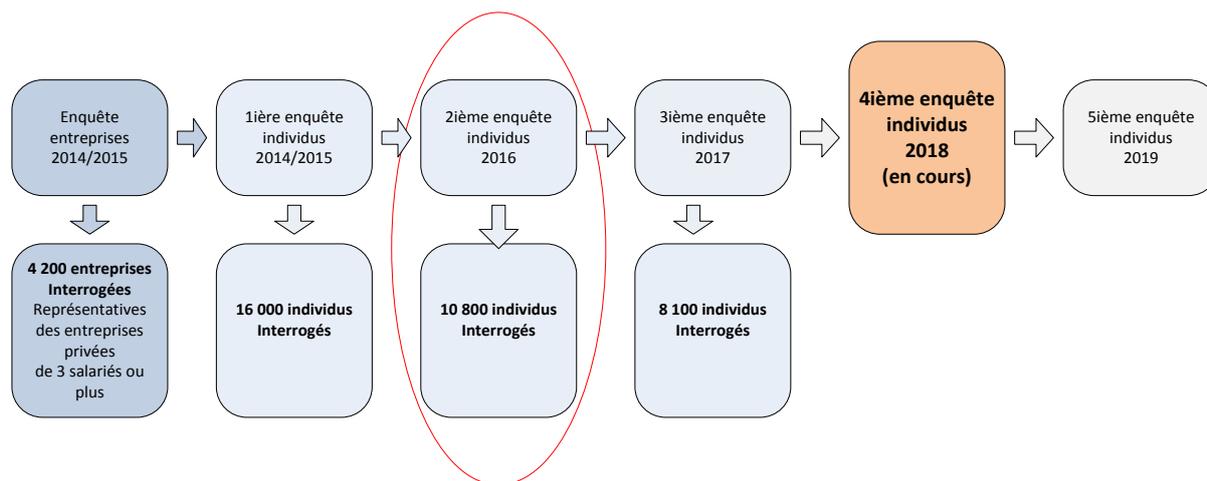
1 - Présentation des données et explication de l'affectation aux différents modes de collectes

Le dispositif d'enquêtes sur les formations et les itinéraires des salariés (Defis) met en relation les pratiques des entreprises avec les parcours professionnels de leurs salariés, ainsi que les formations qu'ils ont suivies. Defis est initié par le Conseil national d'évaluations de la formation professionnelle (CNEFP), financé par le Fonds paritaire de sécurisation des parcours professionnels (FPSPP), conçu par le Céreq et copiloté par le CNEFP et le Céreq.

Ce dispositif est conduit en deux temps :

- un volet « Entreprises » : 4 500 entreprises représentatives du secteur marchand (hors agriculture) à partir de 3 salariés ont été interrogées en 2015,
- un volet « Salariés » : un échantillon de 16 000 individus, salariés en décembre 2013 d'une des entreprises répondantes du volet « Entreprises », a été interrogé une première fois à l'été 2015. Un suivi est prévu pour une durée de 5 ans (jusqu'en 2019).

Les analyses présentées dans cet article s'appuient sur les données issues de la deuxième vague de l'enquête « Salariés » conduite en 2016 auprès de 10 800 individus



La vague 1 du dispositif DEFIS était principalement une collecte monomode par téléphone. En effet le mode de collecte par internet a été uniquement réservé à des individus ne souhaitant pas répondre par téléphone ou à l'enquête, ce qui a abouti à obtenir environ 10% de réponses par internet.

Par contre à partir de la vague 2, l'enquête est devenue véritablement multimode avec un mode de réponse par internet prioritaire dans le processus de collecte. Au final la moitié des individus ont répondu par internet à cette vague.

Afin de pouvoir mesurer lors de cette 2^{ème} vague d'éventuels biais liés au mode de collecte, un échantillon a été aléatoirement tiré parmi les individus destinés à être réinterrogés (c'est à dire la quasi-totalité des répondants à la première vague). L'objectif est alors d'imposer un mode de collecte à 4 sous-groupes de cet échantillon randomisé (groupes correspondant au croisement mode de collecte / adresse mail connue ou non).

Pour être éligibles au tirage, les individus devaient obligatoirement posséder un numéro de téléphone afin d'être contactés dans le cadre de l'enquête par téléphone, ou d'être éventuellement relancés par téléphone pour ceux n'ayant pas répondu par internet (quand ce mode de collecte leur avait été attribué) :

Les individus devant répondre par internet et ne possédant pas d'adresse mail n'ont pas fait l'objet pas de tirage spécifique car ils suivaient en fait un protocole de collecte semblable à certains individus de l'échantillon dit « standard » (c'est-à-dire non randomisé). En effet tous les individus de l'échantillon standard sans mail connu et ayant répondu à l'enquête par internet pourraient alors a posteriori intégrer ce sous-ensemble de l'échantillon permettant des analyses sur les biais liés au mode de collecte.

Le tirage de l'échantillon randomisé a été le suivant :

Echantillon randomisé	Individus avec mail dans le fichier	Individus sans mail dans le fichier
Mode collecte attribué : téléphone	450	450
Mode collecte attribué : internet	1350	Issu de l'échantillon standard

Le volume tiré pour les individus devant répondre par internet est plus élevé car le taux de réponse attendu était plus faible que pour les individus répondant par téléphone, ce qui n'a été finalement pas le cas (voir tableau ci-dessous)

Il est important de souligner que pour cet échantillon randomisé, après une phase visant à une réponse « en première intention » c'est-à-dire dans le mode de collecte assigné, une deuxième phase visait à rattraper par l'autre mode de collecte les non-répondants de cette première phase, de manière à minimiser la non-réponse à l'enquête (ce sous-échantillon randomisé faisant partie de l'échantillon total à interroger).

La structure de l'échantillon randomisé ayant répondu en première intention au mode de collecte attribué a été la suivante :

Echantillon randomisé ayant répondu au mode de collecte attribué (dit de « première intention »)	Individus avec mail connu	Individus sans mail connu	Total
Mode de collecte attribué : téléphone	243	268	511
Mode de collecte attribué : internet	658	939	1597

2- Analyse sur l'échantillon des répondants en première intention :

Les réponses à l'enquête ont donc été collectées par téléphone ou par internet, à part à peu près égale. La question posée lors de l'exploitation de la base de données constituée à l'issue de l'enquête est alors celle de la validité d'une agrégation de ces réponses. Si le mode d'interrogation influence les réponses, il pourrait en effet être judicieux d'en tenir compte dans les études réalisées : peut-on travailler sur ces variables sans corriger le biais de mesure, et quelle est l'influence que ce biais peut avoir sur les analyses ?

L'objectif des deux prochaines parties est de mesurer et d'isoler cet éventuel biais de mesure lié au mode de collecte. On considère généralement que ce biais est compris dans l'effet de mode :

Effet de mode = biais de sélection + biais de mesure

Le biais de sélection signifie que les deux groupes interrogés ne sont pas similaires du point de vue des variables sociodémographiques (X). Dans notre cas, il s'agit de savoir s'il y a une différence dans la distribution des X entre les répondants CATI (Computer Assisted Telephone Interview) et les répondants CAWI (Computer Assisted Web Interview). En d'autres termes, la structure de populations entre nos répondants CATI et CAWI est-elle la même ?

Le biais de mesure signifie que la distribution des Y (variable d'intérêt) conditionnellement à X diffère dans les deux groupes (CATI et CAWI). En d'autres termes les réponses des deux groupes sont-elles différentes dans un mode et dans l'autre, indépendamment des caractéristiques X ?

La différence brute observée entre les réponses des deux groupes constitue l'effet de mode. Pour quantifier le biais de mesure nous devons donc le dissocier du biais de sélection.

Celui-ci est dû à la différence de structure des deux sous populations, par exemple les individus plus jeunes auront une tendance à préférer répondre par internet, alors que les plus âgés seront plus à l'aise avec le téléphone.

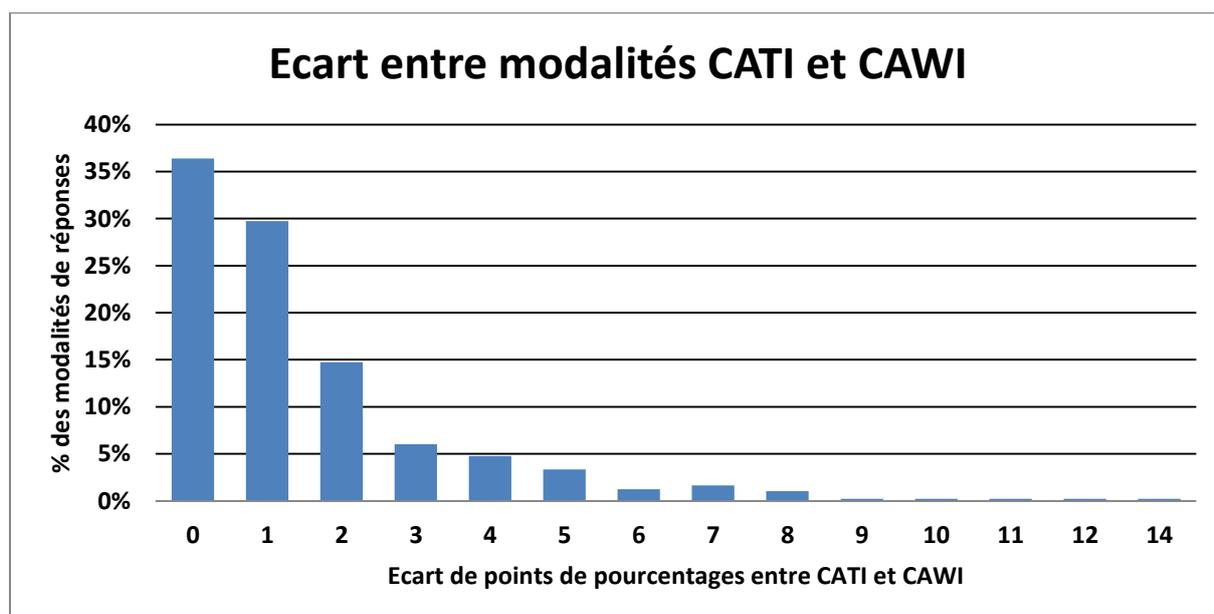
De son côté, l'effet de mesure est dû au fait qu'un même individu ne répondra pas forcément de la même manière devant un questionnaire administré par téléphone ou par internet. Par exemple par téléphone un télé-enquêteur peut expliquer une question mal comprise. Ou alors un individu pourra avoir tendance à donner des réponses plus acceptables socialement par téléphone, ayant peur d'être « jugé » par le télé-enquêteur, alors qu'il pourra apporter des réponses plus sincères par internet.

Dans cette partie nous nous intéressons aux répondants en première intention, c'est-à-dire aux individus ayant répondu dans le mode auquel ils ont été affectés au départ ; en sont donc exclus les

individus répondant dans le mode alternatif, lors de la deuxième phase de collecte. 1597 individus ont été interrogés en première intention par internet et 511 par téléphone. L'affectation « randomisée » permet de construire des groupes comparables sur les caractéristiques personnelles qui auraient pu influencer le fait de répondre par internet ou téléphone s'ils avaient eu le choix, y compris des caractéristiques qui ne sont pas observées dans l'enquête comme l'équipement en outils informatiques, l'aisance dans l'usage d'internet, etc. Cependant, cette « randomisation » n'est pas parfaite, car la propension des individus à répondre à l'enquête peut être différente selon qu'ils sont assignés à un groupe ou un autre. Ainsi, des individus du groupe CATI (resp. CAWI) qui n'ont pas répondu par téléphone (resp. internet) dans cette première phase de collecte en « première intention » ont été invités à répondre dans une deuxième phase par internet (resp. téléphone), de manière à limiter l'attrition. Un effet de sélection demeure donc potentiellement, ce qui nous conduira à tenir compte des caractéristiques sociodémographiques potentiellement différentes dans les deux groupes de répondants, malgré l'assignation aléatoire initiale.

Nous cherchons donc à comparer les réponses des individus répondant par internet (CAWI) et ceux répondants par téléphone (CATI). Pour cela, nous regardons tout d'abord les écarts bruts qui existent entre les deux modes de collecte sur l'échantillon des répondants en première intention. Cette première analyse est réalisée sur l'ensemble des modalités de réponses de l'enquête (environ 600 modalités).

2.1 Mesure des écarts bruts : les effets de modes dans l'enquête Defis



Pour 37% des modalités, il n'existe aucun écart entre les réponses CATI et CAWI, pour 77% des modalités, l'écart est inférieur ou égal à 3 points de pourcentages. Cependant, pour certaines modalités il existe des écarts importants, ce qui nous conduit à penser qu'il existe un effet de mode. Les questions où les écarts sont les plus forts sont principalement des questions d'opinion ou faisant appel à la subjectivité des individus sur lesquelles nous concentrerons notre analyse. Nous savons que ces questions d'opinion et faisant appel à la subjectivité des individus ont tendance à entraîner des biais de « désirabilité sociale », c'est-à-dire la tendance de l'individu à vouloir se montrer sous un jour favorable (Crowne et Marlowe, 1960). Deux facteurs ont été identifiés comme influençant le biais de désirabilité : la nature des questions et la présence de télé-enquêteur (Nass C., Moon Y. et Carney P., 1999).

C'est pour cela que nous nous concentrons sur les variables contenant les expressions « diriez-vous

que » et « pensez-vous que ». 9 variables comportant 24 modalités de réponse sont dans ce cas. Nous effectuons une comparaison des fréquences de ces questions suivie d'un test du Khi-deux. Le test du khi-deux se trouve être significatif pour l'ensemble des variables testées. On rejette donc l'hypothèse d'indépendances des variables d'opinions par rapport à la variable de mode de collecte.

B105a: Dans les 12 prochains mois, pensez-vous avoir un risque de perdre votre emploi?	CATI	CAWI	différence
très probable	3,99	3,64	0,35
probable	7,98	12,98	-5
peu probable	64,79	70,33	-5,54
impossible	23,24	13,05	10,19

Khi-deux<0.0001

B105b: Dans les 12 prochains mois, pensez-vous avoir des chances de promotion?	CATI	CAWI	différence
très probable	1,64	2,03	-0,39
probable	19,01	16,35	2,66
peu probable	52,35	61,44	-9,09
impossible	27	20,18	6,82

Khi-deux<0.0011

B105c: Dans les 12 prochains mois, pensez-vous avoir des chances d'augmentation de salaire?	CATI	CAWI	différence
très probable	3,05	3,94	-0,89
probable	34,27	26,64	7,63
peu probable	44,6	55,7	-11,1
impossible	18,08	13,72	4,36

Khi-deux<0.0001

A307: Diriez-vous que votre niveau de responsabilité?	CATI	CAWI	différence
augmenté	28,7	31,5	-2,8
baissé	3,04	5,12	-2,08
n'a pas évolué	68,26	63,38	4,88

Khi-deux<0.0428

A308: Diriez-vous que votre autonomie ?	CATI	CAWI	différence
augmenté	33,91	37,62	-3,71
baissé	3,7	6,1	-2,4
n'a pas évolué	62,39	56,29	6,1

Khi-deux<0.0146

B104a: Actuellement, diriez-vous que votre travail est intéressant ?	CATI	CAWI	différence
Pas du tout d'accord	47,42	32,27	15,15
Plutôt pas d'accord	41,78	52,22	-10,44
Plutôt d'accord	6,57	11,69	-5,12
Tout à fait d'accord	4,23	3,81	0,42

Khi-deux<0.0001

B104b: Depuis XXXX, diriez-vous que?	CATI	CAWI	différence
Le travail est plus intéressant	28,87	30,47	-1,6
Le travail est moins intéressant	8,22	13,47	-5,25
Il n'y a pas eu de changement dans le travail	62,91	56,06	6,85

Khi-deux<0.0027

B104c: Actuellement, diriez-vous que vos conditions de travail sont pénibles ?	CATI	CAWI	différence
Pas du tout d'accord	14,08	9,53	4,55
Plutôt pas d'accord	27	28,48	-1,48
Plutôt d'accord	31,69	39,73	-8,04
Tout à fait d'accord	27,23	22,26	4,97

Khi-deux<0.0002

A316: Pour votre travail, pensez-vous que ce(s) changement(s) ont été :	CATI	CAWI	différence
Plutôt positifs	14,08	9,53	4,55
Plutôt négatifs	27	28,48	-1,48
Ni l'un ni l'autre	31,69	39,73	-8,04
Pas d'effet	27,23	22,26	4,97

Khi-deux<0.0001

Sur l'ensemble de ces variables, on remarque que la dernière modalité est toujours plus fréquente par téléphone que par internet. Il est possible que, par téléphone, les individus répondent plus facilement à la dernière modalité d'une liste. Cette tendance est appelée « recency effect ».

2.2. Décomposition de l'effet de mode

L'effet de mode est donc avéré, la distribution des Y diffère dans les deux groupes CATI et CAWI. Cependant, les caractéristiques sociodémographiques des deux groupes ne sont pas encore prises en compte. Il faut maintenant séparer biais de sélection et biais de mesure, puisque, comme on l'a vu, si les échantillons CATI et CAWI sont aléatoires, les répondants ne le sont probablement pas.

Dans un premier temps, nous recherchons à identifier des variables sociodémographiques et de proximité avec les outils internet ou téléphone, qui pourraient influencer le fait de répondre par internet ou par téléphone, mais qui ne proviennent pas de cette enquête, de manière à s'assurer qu'aucun biais de mode n'affecte ces variables de contrôle. DEFIS offre de ce point de vue une opportunité intéressante, puisque les caractéristiques sociodémographiques sont collectées lors de la première vague de l'enquête. Cette première vague de DEFIS salariés était principalement une collecte monomode par téléphone. Pour que les variables soient bien issues d'un même mode de collecte, nous écartons les individus interrogés en vague 1 par internet de notre analyse et conservons uniquement les variables provenant du mode de collecte téléphone. Nous utilisons aussi des variables issues de données administratives.

Les variables sociodémographiques sont les suivantes : le sexe, l'âge, le plus haut diplôme, la profession, le fait d'être en couple, le fait d'avoir au moins un enfant, le lieu de naissance, le code postal de la commune de résidence, l'utilisation d'un ordinateur qui influence l'accès à Internet, mais aussi le fait d'avoir passé plus de 50 appels pour joindre l'individu, ce qui indique la plus ou moins grande difficulté de joindre par téléphone. Ces variables sont intégrées dans une régression logistique afin de modéliser leur influence sur le mode de passation. La probabilité modélisée est le fait d'avoir répondu par téléphone (individu traité =1) par rapport à un répondant internet (non traité =0).

$$\text{Logit}[Y(\text{CATI}=1) | X_i] = \beta_0 + \beta_1 * \text{sexe} + \beta_2 * \text{âge} + \beta_3 * \text{diplôme} + \beta_4 * \text{profession} + \beta_5 * \text{couple} + \beta_6 * \text{enfant} + \beta_7 * \text{lieu de naissance} + \beta_8 * \text{ordinateur} + \beta_9 * \text{CP} + \beta_{10} * 50 \text{ appels}$$

Variables	Odds ratio (OR)	Intervalle de confiance (IC) à 95%	P-value	Significativité
SEXE				
Homme (ref)	1			
Femme	0,874	(0,695 - 1,098)	0,2475	NS
Groupes d'âges				
30 ans ou moins	0,634	(0,44 - 0,905)	0,0121	**
31 à 50 ans	1,012	(0,736 - 1,392)	0,9396	NS
plus de 50 ans (ref)	1			
Diplômes				
Aucun diplôme (ref)	1			
diplôme inférieur au BAC	0,829	(0,0,575 - 1,197)	0,3173	NS
BAC général, professionnel ou technologique	0,618	(0,406 - 0,940)	0,0244	**
BAC +2 +3 +4	0,749	(0,496 - 1,132)	0,1705	NS
Bac +5 ou plus	0,592	(0,360 - 0,973)	0,0388	**
Profession déclaré				
Artisans, commerçants et chefs d'entreprise	0,539	(0,227 - 1,279)	0,161	NS
Cadres et professions intellectuels supérieures	0,709	(0,503 - 1,00)	0,0497	**
Professions intermédiaires	0,764	(0,570 - 1,025)	0,073	*
Employés	1,073	(0,795 - 1,449)	0,644	NS
Ouvriers (ref)	1			

Couple				
en couple (ref)	1			
célibataire	0,962	(0,759 - 1,220)	0,7505	NS
Enfants				
a au moins 1 enfant (ref)	1,17			
n'a pas d'enfant	1,17	(0,942 - 1,453)	0,1551	NS
Lieux de naissance				
France (ref)	1			
Etranger	1,624	(1,168 - 2,659)	0,0039	***
Utilise même occasionnellement un ordinateur				
Utilise (ref)	1			
N'utilise pas	1,673	(1,307 - 2,142)	<.0001	***
Code postal validé				
Oui (ref)				
Non	0,935	(0,396 - 2,208)	0,8789	NS
Plus de 50 appels				
Oui (ref)				
Non	0,480	(0,384 - 0,600)	<.0001	***

Note de lecture : Un individu n'utilisant pas un ordinateur a 1.6 fois plus de chance de répondre par téléphone que par internet par rapport à un individu qui en utilise un.

Malgré la randomisation de l'échantillon, il existe bien un effet de sélection entre les répondants CATI et CAWI. En effet, la régression logistique montre que les probabilités d'appartenir à un groupe ou à l'autre dépendent en partie de caractéristiques individuelles des individus comme l'âge, le niveau de diplôme, le lieu de naissance, l'utilisation d'un ordinateur ou le nombre d'appels téléphoniques. Cependant la significativité des variables de la régression ci-dessus n'est pas très élevée, notamment si on la compare avec la même régression effectuée sur l'ensemble des répondants (voir ci-dessous).

Pour savoir si la distribution des Y (nos variables d'opinions) conditionnellement aux X (les variables sociodémographiques) diffère chez les répondants CATI et CAWI, nous effectuons une série de régressions logistiques, où nous estimons la probabilité d'avoir répondu à une des modalités de nos variables d'opinions en fonctions des variables sociodémographiques et de la variables CATI/CAWI.

$$\text{Logit}[Y(\text{Variables d'opinions}=1) | X_i] = \beta_0 + \beta_1 * \text{sexe} + \beta_2 * \text{âge} + \beta_3 * \text{diplôme} + \beta_4 * \text{profession} + \beta_5 * \text{couple} + \beta_6 * \text{enfant} + \beta_7 * \text{lieu de naissance} + \beta_8 * \text{ordinateur} + \beta_9 * \text{CP} + \beta_{10} * 50 \text{ appels} + \beta_{11} * \text{CATI}$$

Le tableau suivant illustre la significativité de la variable CATI/CAWI pour chaque question d'opinion préalablement sélectionnée.

Numéro de la question	Question:	Modalités	Pr < khi-2 de la variable CATI/CAWI logit multivarié	Seuil multivarié
B105a	Dans les 12 prochains mois, pensez-vous avoir un risque de perdre de votre emploi?	1- très probable 2- probable 3- peu probable 4- impossible	<.0001	***

B105b	Dans les 12 prochains mois, pensez-vous avoir des chances de promotion?	1- très probable 2- probable 3- peu probable 4- impossible	0,8902	NS
B105c	Dans les 12 prochains mois, pensez-vous avoir des chances d'augmentation de salaire?	1- très probable 2- probable 3- peu probable 4- impossible	0,0119	**
A307	Diriez-vous que votre niveau de responsabilité?	1- augmenté 2- baissé 3- n'a pas évolué	0,5747	NS
A308	Diriez-vous que votre autonomie ?	1- augmenté 2- baissé 3- n'a pas évolué	0,3438	NS
A316	Pour votre travail, pensez-vous que ce(s) changement(s) ont été :	1- plutôt positifs 2- plutôt négatif 3- ni l'un ni l'autre 4- pas d'effet	0,0032	***
B104a	Actuellement, diriez-vous que votre travail est intéressant ?	1- Tout à fait d'accord 2- Plutôt d'accord 3- Plutôt pas d'accord 4- Pas du tout d'accord	<.0001	***
B104b	Depuis XXXX, diriez-vous que?	1- Votre travail est plus intéressant 2- Votre travail est moins intéressant 3- Il n'y a pas eu de changement dans votre travail	0,3382	NS
B104c	Actuellement, diriez-vous que vos conditions de travail sont pénibles ?	1- Tout à fait d'accord 2- Plutôt d'accord 3- Plutôt pas d'accord 4- Pas du tout d'accord	0,0135	**

B104d	Depuis XXXX, diriez-vous que ?	1- Vos conditions de travail se sont améliorées 2- Vos conditions de travail se sont détériorées 3- Elles n'ont pas changé	0,1581	NS
-------	--------------------------------	--	--------	----

Sur les dix variables d'opinions testées, la variable CATI/CAWI est significative au seuil de 5% au minimum dans cinq cas. Sous réserve qu'il n'y a pas de variables sociodémographiques inobservées qui auraient une influence significative sur la sélection des répondants en CATI vs CAWI de l'échantillon randomisé, le fait de répondre par internet ou par téléphone influence la réponse que l'individu fera à la moitié des variables d'opinions.

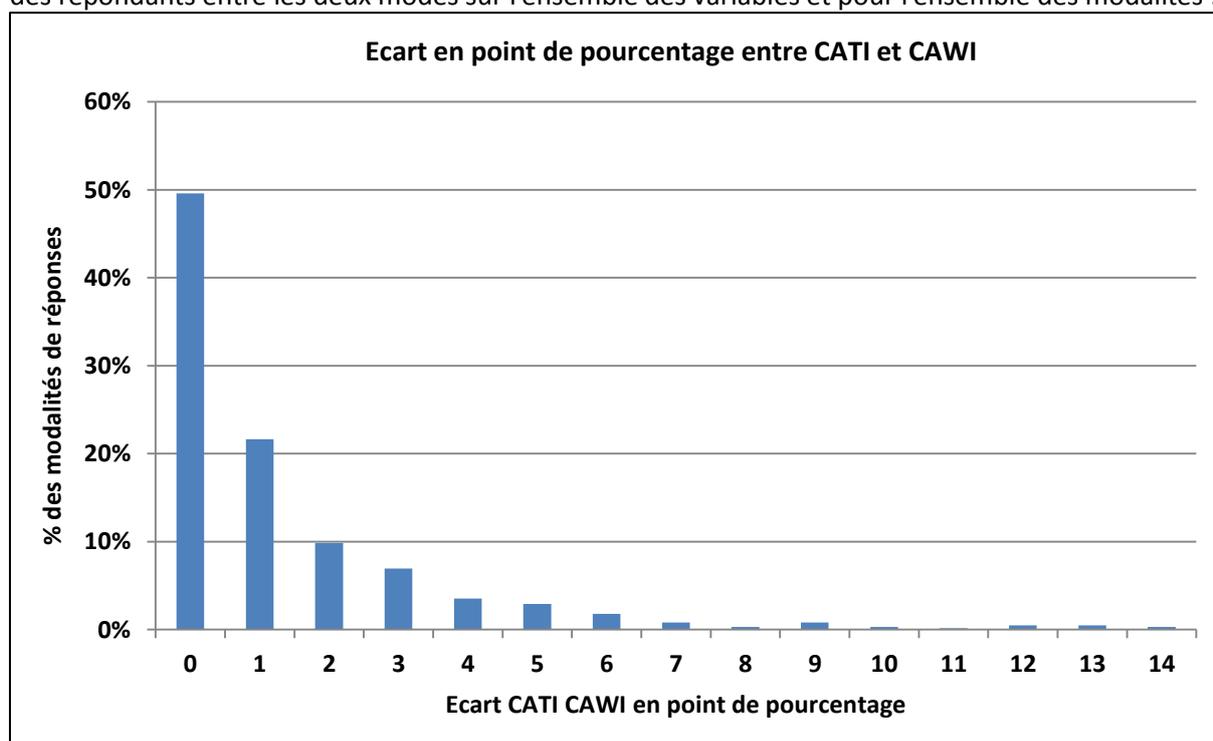
A ce stade, on observe que **la distribution des Y conditionnellement aux X diffère en CATI et en CAWI ; il existe alors un effet de mesure.**

Cependant, puisqu'on a montré que sur l'échantillon randomisé un effet résiduel de sélection des répondants pendant la phase dédiée à l'expérimentation est observé, pourquoi se restreindre à celui-ci ?

Nous allons donc élargir notre champ d'investigation à l'ensemble des répondants de l'enquête et essayer d'annuler le biais de sélection par appariement sur score de propension afin d'isoler l'effet de mode et de le quantifier pour ces variables d'opinion.

3- Analyse sur l'ensemble de l'échantillon répondant :

Cette fois sur l'ensemble des répondants, nous reprenons la quantification de l'écart brut de réponse des répondants entre les deux modes sur l'ensemble des variables et pour l'ensemble des modalités :



Pour près de 50 % des modalités de réponse, les répondants par téléphone répondent de la même manière que les répondants par internet. Pour 22 % des modalités, il existe une différence de 1 point de pourcentage entre les deux modes.

Comme en deuxième partie, nous modélisons la probabilité d'avoir répondu par téléphone (individu traité =1) par rapport à un répondant téléphone (non traité =0), en retenant les mêmes variables de contrôle.

$$\text{Logit}[Y(\text{CATI}=1) | X_i] = \beta_0 + \beta_1 \cdot \text{sexe} + \beta_2 \cdot \text{âge} + \beta_3 \cdot \text{diplôme} + \beta_4 \cdot \text{profession} + \beta_5 \cdot \text{couple} + \beta_6 \cdot \text{enfant} + \beta_7 \cdot \text{lieu de naissance} + \beta_8 \cdot \text{ordinateur} + \beta_9 \cdot \text{CP} + \beta_{10} \cdot \text{50 appels}$$

Variables	Odds ration (OR)	Intervalle de confiance (IC) à 95%	P-value	Significativité
SEXE				
Homme (ref)	1			
Femme	0,821	(0,746 - 0,904)		***
Groupes d'âges				
30 ans ou moins	0,766	(0,660 - 0,888)	0,0004	***
31 à 50 ans	0,728	(0,633 - 0,837)	<.0001	***
plus de 50 ans (ref)	1			
Diplômes				
Aucun diplôme (ref)	1,000			
diplôme inférieur au BAC	0,694	(0,582 - 0,827)	<.0001	***
BAC général, professionnel ou technologique	0,573	(0,473 - 0,694)	<.0001	***
BAC +2 +3 +4	0,520	(0,430 - 0,630)	<.0001	***
Bac +5 ou plus	0,471	(0,379 - 0,587)	<.0001	***
Profession déclaré				
Artisans, commerçants et chefs d'entreprise	0,776	(0,544 - 1,055)	0,1057	NS
Cadres et professions intellectuels supérieures	0,685	(0,593 - 0,792)	<.0001	***
Professions intermédiaires	0,71	(0,627 - 0,806)	<.0001	***
Employés	0,899	(0,787 - 1,027)	0,117	NS
Ouvriers (ref)	1			
Couple				
en couple (ref)				
célibataire	1,123	(1,014 - 1,244)	0,0254	**
Enfants				
a au moins 1 enfant (ref)	1			
n'a pas d'enfant	0,91	(0,828 - 1,000)	0,0490	*
Lieux de naissance				
France (ref)	1			
Etranger	1,28	(1,088 - 1,507)	0,0029	***
Utilise même occasionnellement un ordinateur				
Utilise (ref)	1			
N'utilise pas	1,606	(1,438 - 1,793)	<.0001	***
Code postal validé				
Oui (ref)	1			
Non	3,090	(2,316 - 4,123)	<.0001	***
Plus de 50 appels pour qualifier				
Oui (ref)	1			
Non	0,458	(0,413 - 0,507)	<.0001	***

Le fait d'être une femme diminue les chances d'avoir répondu par téléphone de 82.1 % au seuil de 1%. Le fait d'avoir moins de 30 ans diminue les chances de répondre par téléphone de 76.6 % par rapport au fait d'avoir plus de 50 ans au seuil de 1% etc...

On remarque que par rapport à l'échantillon randomisé, ces variables sont beaucoup plus significatives, ce qui laisse à penser que le biais de sélection et/ou le biais de mesure est plus important. Cependant, s'il est logique que le biais de sélection soit plus important lorsqu'on laisse dès le début la possibilité à ceux qui le préfèrent d'utiliser internet, il est difficile d'imaginer la raison d'un biais de mesure plus important pour l'ensemble des répondants à l'enquête par rapport à ceux de l'échantillon randomisé.

Pour minimiser au maximum l'effet de sélection, et quantifier l'effet de mesure, nous utilisons une méthode de matching en **considérant le mode de collecte comme un traitement** et ainsi nous rapprocher de méthodes utilisées dans l'évaluation des politiques publiques.

Ce matching se fera à partir **du score de propension**. Le score de propension est la probabilité d'être exposé à un traitement en fonction de caractéristiques individuelles. Cette probabilité est comprise entre 0 et 1. Nous calculerons le score de propension à partir d'une régression logistique.

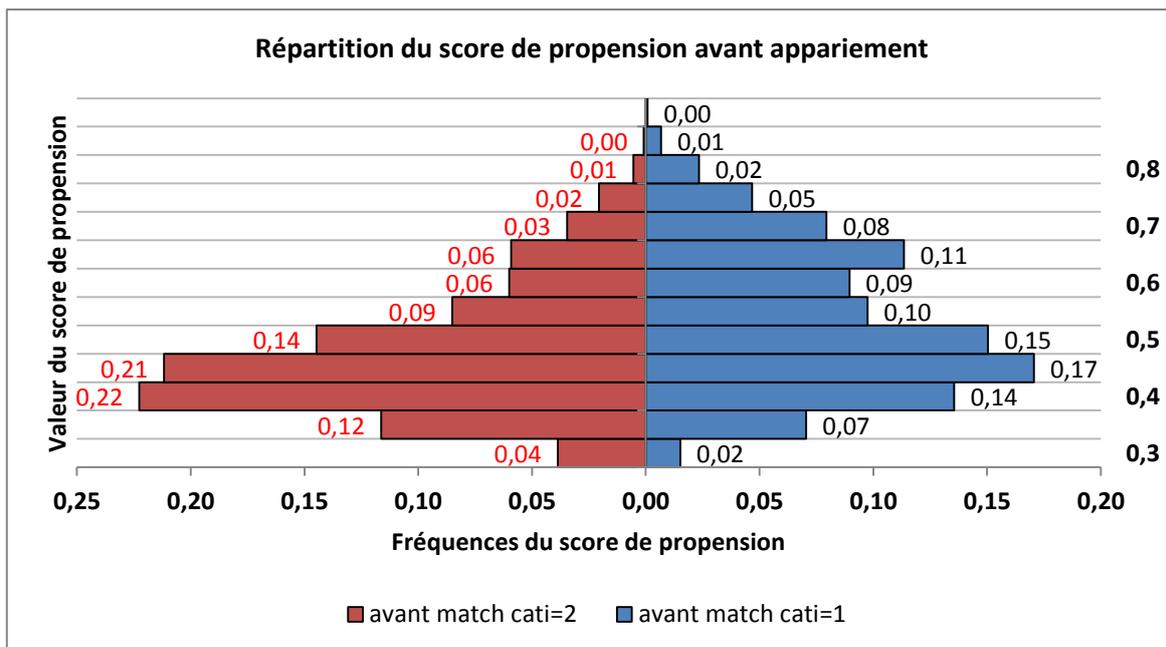
$e(x_i)$ estime pour chaque individu i sa probabilité P de répondre dans un mode ou l'autre Z , en fonction des variables sociodémographiques intégrés dans la régression (x_i)

$$e(x_i) = P(z_i = 1 | x_i)$$

Deux hypothèses sont nécessaires pour effectuer un score de propension :

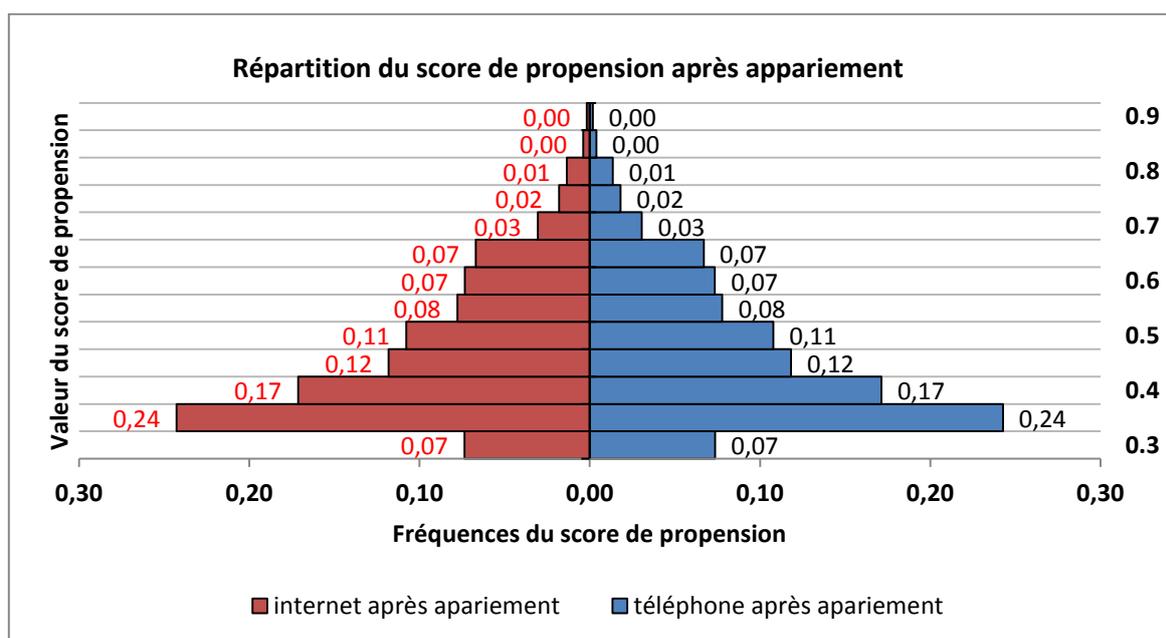
- **l'hypothèse d'indépendance conditionnelle à des caractéristiques observables**
Cette hypothèse signifie que l'on peut réduire le biais de sélection s'il existe des variables indépendantes de l'affectation du traitement. Dans notre cas les variables sociodémographiques utilisées proviennent soit de la vague 1 et d'un seul mode de collecte (répondants téléphone) soit de données administratives. Dans les deux cas elles sont indépendantes du traitement (le mode de collecte).
- **L'hypothèse de la condition du support commun**
Cette hypothèse signifie que la répartition du score de propension dans le groupe traité et non traité permet de trouver des individus comparables dans chaque groupe. C'est-à-dire qu'il existe une zone commune de la distribution de la probabilité d'être traité (c'est-à-dire d'être affecté à un mode de collecte)

Les individus ayant leurs variables sociodémographiques similaires auront des scores de propension très proches. Deux individus ayant le même score de propension, (chacun ayant répondu dans un mode différent), pourront être appelés contrefactuel. Pour avoir un aperçu de la symétrie de notre score de propension garant d'une structure équilibrée des deux sous-populations (CATI et CAWI) nous comparons la répartition des deux scores.



On constate que la répartition du score de propension avant appariement chez les répondants CAWI et CATI n'est pas symétrique, ce qui illustre le biais de sélection entre les deux populations de répondants.

Le support commun des scores de propensions est compris entre le minimum le plus élevé du CATI ou du CAWI et le maximum le plus faible du CATI et du CAWI. Le support commun permet d'éclaircir la qualité du score propension. Ici la zone de support commun est comprise entre 0.226032 et 0.9235880. Nous éliminons les individus hors support commun car ils ne trouveraient pas de contrefactuel dans le groupe opposé. Le support commun étant étendu (80% des individus ont un score de propension qui appartient au support commun), nous effectuerons un appariement sur score de propension par la méthode du plus proche voisin et avec un caliper de 0.2, le caliper étant la distance maximum entre deux paires d'individus appariées. Avant appariement, nous avons 5046 individus répondant en CATI et 5525 répondants en CAWI. Après l'appariement nous avons 4244 répondants en CATI et 4244 répondants en CAWI. 2083 individus ont été supprimés pour réduire l'effet de sélection. Après l'appariement, la répartition du score de propension est la suivante :



Cette nouvelle répartition du score de propension après appariement est beaucoup plus équilibrée. Il y a une symétrie entre CATI et CAWI. Nous regardons désormais de combien notre appariement a réduit le biais de sélection sur nos variables sociodémographiques.

Variables	Echantillons	Mode		Biais		t-test	
		CATI	CAWI	% Biais	% de réduction du biais	t	P< t
30 ans ou moins	non apparié	0,325	0,334	1,91%	44%	-1,0	0,30
	apparié	0,328	0,323	-1,06%		0,5	0,626
31 ans à 50 ans	non apparié	0,55	0,571	4,20%	81%	-2,1	0,033
	apparié	0,558	0,562	0,80%		-0,4	0,693
Plus de 50 ans	non apparié	0,125	0,095	-9,68%	100%	5,0	<0,0001
	apparié	0,115	0,115	0,00%		-0,1	0,918
lieu de naissance	non apparié	0,086	0,054	-12,80%	78%	6,6	<0,0001
	apparié	0,072	0,065	-2,80%		1,3	0,196
Couple	non apparié	0,28	0,25	-6,82%	0%	2,9	0,0034
	apparié	0,3	0,27	-6,82%		2,9	0,0035
Enfant	non apparié	0,471	0,452	-3,80%	53%	2,0	0,0464
	apparié	0,473	0,464	-1,80%		0,8	0,409
Ordinateur	non apparié	0,331	0,172	-36,98%	82%	19,3	<0,0001
	apparié	0,251	0,223	-6,51%		30,6	0,0022
Code postal validé	non apparié	0,962	0,988	18,57%	62%	-8,5	<0,0001
	apparié	0,974	0,984	7,14%		-3,1	0,002
Plus de 50 appels	non apparié	0,261	0,14	-30,25%	73%	15,8	<0,0001
	apparié	0,216	0,183	-8,25%		3,8	0,002
Sexe	non apparié	0,279	0,324	9,78%	71%	-5,1	<0,0001
	apparié	0,301	0,288	-2,83%		1,3	0,207
Artisans, commerçants et chefs d'entreprise	non apparié	0,017	0,02	2,14%	100%	-1,1	0,257
	apparié	0,02	0,02	0,00%		0,0	1
Cadres et professions intellectuelles supérieures	non apparié	0,202	0,301	22,50%	91%	-	<0,0001
	apparié	0,23	0,239	2,05%		-1,0	0,3435
Professions intermédiaires	non apparié	0,221	0,279	13,49%	97%	-7,0	<0,0001
	apparié	0,252	0,254	0,47%		-0,2	0,8417
Employés	non apparié	0,193	0,18	-3,33%	85%	1,8	0,0797
	apparié	0,201	0,203	0,51%		-0,2	0,8078
Ouvriers	non apparié	0,367	0,219	-32,89%	90%	16,9	<0,0001
	apparié	0,299	0,284	-3,33%		1,3	0,1969
Aucun diplôme	non apparié	0,097	0,043	-21,72%	65%	11,2	<0,0001
	apparié	0,074	0,056	-7,52%		3,5	0,004

Brevet, CAP ou BEP	non apparié	0,363	0,264	-21,52%	96%	11,0	<0,0001
	apparié	0,331	0,335	0,87%		-0,4	0,6617
Bac (tous types)	non apparié	0,179	0,184	1,28%	-60%	-0,7	0,518
	apparié	0,189	0,197	2,05%		-0,9	0,3795
Bac +2 à Bac +4	non apparié	0,244	0,324	17,78%	85%	-9,1	<0,0001
	apparié	0,271	0,283	2,67%		-1,2	0,2164
Bac +5 et plus	non apparié	0,115	0,194	22,07%	94%	-1,0	<0,0001
	apparié	0,133	0,128	-1,40%		0,7	0,4788

Ce tableau nous présente la réduction du biais de sélection pour chaque variable du modèle et nous permet de confirmer ce que nous avons constaté graphiquement avec la nouvelle répartition du score de propension. Nous comparons les fréquences et les moyennes avant et après appariement entre répondants CATI et CAWI et regardons la significativité de la différence des moyennes ou fréquences par un t-test. Le calcul du biais s'obtient en divisant la différence des moyennes CATI et CAWI par l'écart type commun de l'échantillon. Cette opération est effectuée avant et après appariement, ce qui nous permet d'obtenir un taux de réduction du biais par différence entre le biais avant appariement et celui après appariement.

Ainsi, pour l'ensemble des variables le biais est réduit à l'exception de la variable diplôme (modalité « tous types de BAC »).

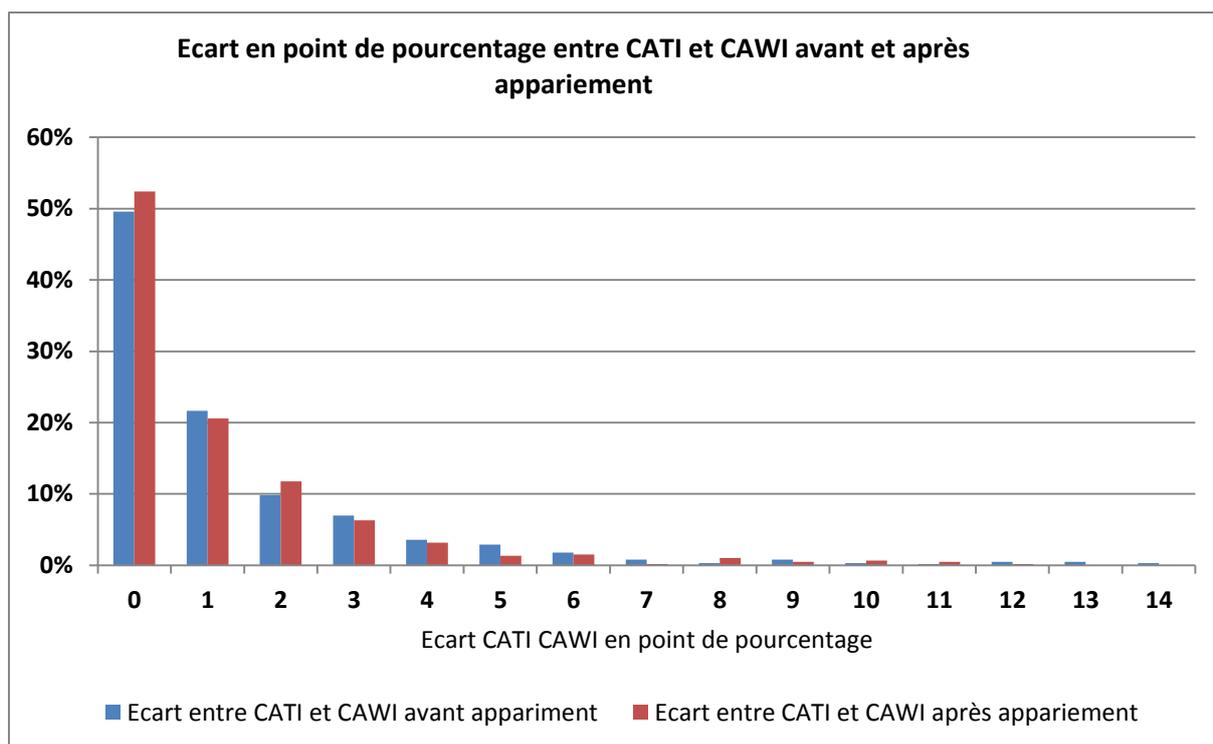
Par exemple :

- Avant appariement, pour la variable lieu de naissance 8.6% des individus sont nés à l'étranger chez les répondants CATI et 5.4% chez les répondants CAWI : cet écart est significatif selon le test de Student et le biais est de 12.8%.
- Après appariement, 7.2% des répondants CATI et 6.5% des répondants CAWI sont nés à l'étranger, cet écart n'est plus significatif. Le biais est désormais de 2.8%. Il a donc été réduit grâce à l'appariement de 78%.

Nous nous intéressons de nouveau à la différence de fréquences de chaque modalité entre les répondants CATI et CAWI

Différence CATI CAWI avant appariement					
Minimum	Quartile inférieur	Médiane	Moyenne	Quartile supérieur	Maximum
0	0	1,0	1,388	2,0	14,0
Différence CATI CAWI après appariement					
Minimum	Quartile inférieur	Médiane	Moyenne	Quartile supérieur	Maximum
0	0	0	1,196	2,0	13,0

On remarque une diminution de la différence moyenne, ainsi que de la médiane et du maximum.



Dans l'échantillon apparié, il y a pour désormais plus de 52% des modalités pour lesquelles il n'existe pas d'écart entre les répondants CATI et CAWI.

Nous mettons en œuvre une série de régression logistique sur les variables d'opinions contenant les mêmes variables explicatives que précédemment et nous rajoutons la variable CATI/CAWI afin de connaître sa significativité. Si la variable CATI/CAWI est significative en l'absence de biais de sélection nous sommes alors en présence d'un biais de mesure.

$$\text{Logit}[Y(\text{CATI}=1) | X_i] = \beta_0 + \beta_1 \cdot \text{sexe} + \beta_2 \cdot \text{âge} + \beta_3 \cdot \text{diplôme} + \beta_4 \cdot \text{profession} + \beta_5 \cdot \text{couple} + \beta_6 \cdot \text{enfant} + \beta_7 \cdot \text{lieu de naissance} + \beta_8 \cdot \text{ordinateur} + \beta_9 \cdot \text{CP} + \beta_{10} \cdot 50 \text{ appels}$$

Numéro de la question	Question:	Modalités	Pr < khi-2 de la variable CATI/CAWI logit multivarié	Seuil significativité
B105a	Dans les 12 prochains mois, pensez-vous avoir un risque de perdre de votre emploi?	1- très probable 2- probable 3- peu probable 4- impossible	<.0001	***
B105b	Dans les 12 prochains mois, pensez-vous avoir des chances de promotion?	1- très probable 2- probable 3- peu probable 4- impossible	<.0001	***

B105c	Dans les 12 prochains mois, pensez-vous avoir des chances d'augmentation de salaire?	1- très probable 2- probable 3- peu probable 4- impossible	<.0001	***
A307	Diriez-vous que votre niveau de responsabilité?	1- augmenté 2- baissé 3- n'a pas évolué	0,2007	NS
A308	Diriez-vous que votre autonomie ?	1- augmenté 2- baissé 3- n'a pas évolué	0,5910	NS
A316	Pour votre travail, pensez-vous que ce(s) changement(s) ont été :	1- plutôt positifs 2- plutôt négatif 3- ni l'un ni l'autre 4- pas d'effet	<.0001	***
B104a	Actuellement, diriez-vous que votre travail est intéressant ?	1- Tout à fait d'accord 2- Plutôt d'accord 3- Plutôt pas d'accord 4- Pas du tout d'accord	<.0001	***
B104b	Depuis XXXX, diriez-vous que?	1- Votre travail est plus intéressant 2- Votre travail est moins intéressant 3- Il n'y a pas eu de changement dans votre travail	0,9369	NS
B104c	Actuellement, diriez-vous que vos conditions de travail sont pénibles ?	1- Tout à fait d'accord 2- Plutôt d'accord 3- Plutôt pas d'accord 4- Pas du tout d'accord	<.0001	***
B104d	Depuis XXXX, diriez-vous que ?	1- Vos conditions de travail se sont améliorées 2- Vos conditions de travail se sont détériorées 3- Elles n'ont pas changé	0,4470	NS

Sur les dix variables d'opinions testées six sont concerné par un biais de mesure. Les questions concernées par ce biais sont les questions comprenant quatre modalités.

Conclusion :

Pour une grande majorité des variables l'effet de mesure est faible ou inexistant. Cependant nous avons constaté que pour certaines variables d'opinions il existe un effet de mesure plus ou moins important, accentué par « recency effect ». A noter que sur ces variables, le biais de mesure, significatif, ne modifie cependant pas le sens de l'analyse. C'est le cas des variables d'opinion présentant des modalités de réponse sous forme d'une échelle en quatre modalités, les deux dernières modalités (libellées selon le cas « plutôt pas d'accord » et « pas du tout d'accord » ; « impossible » et « peu probable ») concentrant l'essentiel de l'écart.

Le fait de regrouper ces modalités, ou plus généralement les modalités extrêmes peut permettre de diminuer sensiblement cet effet.

mode de collecte	Dans les 12 prochains mois, pensez-vous avoir un risque de perdre votre emploi?	
	Oui	Non
Téléphone	17,32	82,68
Internet	19,72	80,28

mode de collecte	Dans les 12 prochains mois, pensez-vous avoir des chances de promotion?	
	Oui	Non
Téléphone	25,02	74,98
Internet	17,45	82,55

mode de collecte	Dans les 12 prochains mois, pensez-vous avoir des chances d'augmentation de salaire?	
	Oui	Non
Téléphone	25,02	74,98
Internet	17,45	82,55

mode de collecte	Actuellement, diriez-vous que vos conditions de travail sont pénibles ?	
	Oui	Non
Téléphone	41,93	58,07
Internet	41,26	58,74

mode de collecte	Actuellement, diriez-vous que votre travail est intéressant ?	
	Oui	Non
Téléphone	89,13	10,87
Internet	83,63	16,37

Nous recommandons ainsi, pour l'analyse, de noter la présence d'un biais de mesure pour ces variables d'opinion en quatre modalités indiquant une échelle d'accord avec la proposition, et de les analyser de manière regroupée en deux modalités.

Après l'étude du biais que nous venons de mener, nous nous interrogeons sur l'utilité de la mise en

place d'un protocole lourd permettant d'obtenir un échantillon de contrôle. En effet, d'une part, le biais de sélection n'a pas été totalement éliminé par ce protocole, car si l'affectation dans les sous-échantillons était bien aléatoire, on a pu observer un biais de sélection résiduel parmi les répondants de chacun des deux groupes. D'autre part, la proportion de répondants dans les deux modes de collecte étant assez similaire, il est possible d'utiliser aisément la méthode d'appariement sur score de propension sans éliminer trop d'individus, et d'obtenir une quantification des biais de mesure qui nous semble robuste.

Bibliographie :

- [1] Christophe BARRET, Mady CISSÉ, « Agrégation de données multimode : impact sur la modélisation des variables présentant un effet de mesure » JMS 2018
- [2] Lecocq, A., Ammi, M. & Bellarbre, É. « Le score de propension : un guide méthodologique pour les recherches expérimentales et quasi expérimentales en éducation » *Mesure et évaluation en éducation*, 37(2), 69–100. 2014.
- [3] Paul R. Rosenbaum and Donald B. Rubin «The Central Role of the Propensity Score in Observational Studies for Causal Effects» *Biometrika* Vol. 70, N° 1 pp. 41-55, Avril 1983
- [4] Christophe Dzikowski, Christophe Barret, « Comparaison des modes de collecte par téléphone et par internet dans le cadre de l'enquête génération du Céreq. Analyse des abandons en cours et tentative pour distinguer les effets de sélection et de mesure »
- [5] Tiaray Razafindranovona, « La collecte multimode et le paradigme de l'erreur d'enquête totale » *Méthodologie statistique*, M 2015/01
- [6] Thompson Mary E, Huang Y. Celia, Boudreau Christian *et al.*, « Mesurer les effets de la méthode de collecte des données. Application à l'enquête *International Tobacco Control* des Pays-Bas », *Population*, 2013/3 (Vol. 68), p. 511-535
- [7] Ying So Warren F. Kuhfeld, « *Multinomial Logit Models* »
- [8] Clifford Nass, Youngme Moon, Paul Carney “*Are People Polite to Computers? Responses to Computer Based Interviewing Systems*” 1999
- [9] Crowne, D. P., & Marlowe, D., “*A new scale of social desirability independent of psychopathology*”. *Journal of Consulting Psychology*, 24(4), 349-354. (1960).
- [10] Raphaëlle Butori, Béatrice Parguel. « *Les biais de réponse - Impact du mode de collecte des données et de l'attractivité de l'enquêteur* ». AFM, 2010, France. 2010.