

COMPARAISON DES MODES DE COLLECTE PAR TELEPHONE ET
PAR INTERNET DANS LE CADRE DE L'ENQUETE GENERATION DU
CEREQ.

ANALYSE DES ABANDONS EN COURS ET TENTATIVE POUR
DISTINGUER LES EFFETS DE SELECTION ET DE MESURE

Christophe Dzikowski¹ (Céreq) & Christophe Barret² (Céreq)

Céreq, 10 Place de la Joliette, BP 21321, 13567 Marseille cedex 02, dzikowski@cereq.fr Auteur 1

Céreq, 10 Place de la Joliette, BP 21321, 13567 Marseille cedex 02, barret@cereq.fr Auteur 2

Résumé

Le dispositif Génération a pour objectif l'étude des parcours d'insertion professionnelle des jeunes débutants. À cet effet un panel de jeunes sortants du système éducatif est interrogé par téléphone à 3, 5, et 7 ans après leur entrée sur le marché du travail. Ce dispositif permet de rendre compte des interactions entre la formation initiale et le parcours professionnel. Une expérimentation de collecte par internet (CAWI : Computer-Assisted Web Interviewing) a été menée en parallèle de l'enquête 2013 auprès de la Génération 2010, sur un échantillon disjoint de celui de la collecte par téléphone (CATI : Computer-Assisted Telephone Interviewing).

Le questionnaire passé sous CAWI est, à quelques modifications marginales, identique au questionnaire CATI. Les nombreux abandons en cours d'enquête sont vraisemblablement liés au manque d'ergonomie du questionnaire web, à sa longueur mais aussi de manière plus marquée à certaines questions difficiles notamment en présence de consignes longues. Le risque d'abandonner est aussi clairement différencié selon le niveau de sortie.

Pour distinguer les effets de sélection des effets de mesure, une première méthode consiste à corriger de la non-réponse et à caler sur des structures identiques les collectes CATI et CAWI. En vue d'interpréter les écarts d'estimations mesurés entre les deux populations, la variance des variables d'intérêt sera estimée par calcul direct. En fonction des écarts entre les estimations sur les deux collectes, des méthodes de matching ont été mobilisées pour essayer de mieux distinguer les effets de sélection des effets de mesures avec toutes les difficultés que cela comporte.

Mots clés .CAWI, CATI, abandons en cours, effets de mode, effet de sélection, effet de mesure, calcul de précision, matching

Introduction

Le Centre d'études et de recherche sur les qualifications (Céreq) a pour mission l'étude de la relation entre la formation et l'emploi ou celui de la formation continue, ou encore la professionnalisation. Le département entrée et évolution dans la Vie Active (Deeva) étudie en particulier les parcours d'insertion des jeunes à la sortie du système éducatif. Pour cela, le département s'appuie essentiellement sur le dispositif d'enquêtes Génération. Les jeunes quittant pour la première fois le système éducatif sont interrogés à 3, 5 et 7 ans après leur sortie. Depuis 1998, cette enquête se fait sous CATI, computer assisted telephone interviewing. Les contraintes budgétaires imposent de repenser l'organisation de cette enquête. Plusieurs pistes sont envisagées : un appariement avec des sources administratives, une réduction du questionnaire, de l'échantillon, voire de la périodicité des enquêtes Générations et un nouveau mode de collecte, reposant pour tout ou partie sur une collecte par internet. A l'initiative du Céreq, une expérimentation internet a été menée en parallèle de l'enquête par téléphone courant 2013. Cette étude a pour objectif d'effectuer une comparaison méthodologique de ces deux modes de collecte.

Cette étude s'attachera à présenter dans un premier temps un bilan de la collecte, à savoir la qualité de la base de sondage et l'intensité des abandons en cours dans la collecte par internet. Une courbe de survie des répondants au cours de la passation du questionnaire sera utilisée pour identifier les types de questions qui posent le plus de difficultés et génèrent le plus d'abandons. Par la suite il s'agira de présenter une méthode pour corriger des effets de sélection. Une comparaison des indicateurs issues des collectes internet et téléphone montrera que pour la majorité des questions des différences significatives persistent. Ces dernières peuvent être dues aux effets de mesure (présence ou non de la modalité «Ne sais pas » par exemple), ou aux effets de sélection qui ne sont que partiellement corrigés. Une troisième partie s'intéressera aux calculs de précisions des estimations. Il s'agit ici d'identifier les questions pour lesquelles les écarts les plus significatifs sont observés. Enfin on présentera une autre méthode pour tenter de mieux isoler les effets de sélection des effets de mesure, la méthode de matching.

1 Bilan de la collecte

La base de sondage des jeunes présumés sortants du système éducatif est construite par le Céreq en collectant l'information auprès de l'ensemble des établissements dispensant une formation initiale. Pour cette enquête, l'email était peu renseigné. On estime à partir du suivi des envois de mails que

16% des sortants peuvent effectivement être contactés par courriel. L'indisponibilité des adresses emails n'est pas homogène suivant les types d'établissements des sortants. Ainsi on estime que 40% des jeunes sortants des universités, des écoles de commerce ou des écoles d'ingénieur sont joignables, alors que pour les jeunes sortants des rectorats nous n'avons pas d'adresses email. Une tentative d'enrichissement des adresses emails a été faite auprès d'un prestataire, mais le résultat n'a pas été concluant. Sur les 264 000 individus de l'échantillon CAWI, 64 000 mails ont été envoyés. 48 000 mails ont effectivement été reçus. 16 000 connexions sur le lien du questionnaire ont été enregistrées et 12 000 répondants ont terminé la première partie du questionnaire qui permet de déterminer les individus appartenant au champ de l'enquête (questionnaire filtre). Parmi eux, 7400 individus sont dans le champ et 4600 sont hors champ. Parmi les individus répondants dans le champ, seuls 4500 jeunes ont terminé le questionnaire, deux jeunes sur 5 ont donc abandonné le questionnaire en cours d'enquête.

Les abandons en cours de passation ne se retrouvent pas avec la même intensité dans la collecte

par téléphone où le taux d'abandon est quatre fois moindre que pour la collecte internet. On retrouve ici le rôle essentiel du télé-enquêteur, formé spécifiquement à l'enquête, chargé d'assumer la complexité du questionnaire, de reformuler les questions en cas d'incompréhension, de maintenir l'enquêté captif. On retrouve le constat souvent établi dans les études méthodologiques d'enquêtés moins impliqués dans les enquêtes auto administrées comme le rappelle Frippiat et Marquis (2010).

En plus de cette moindre implication, deux types d'abandons distincts semblent se produire au cours. Tout d'abord un abandon diffus qui se traduit par une décroissance de la courbe de survie avec une faible pente. Ces abandons diffus pourraient être attribués au manque d'intérêt des questions pour l'enquêté ou encore à la faible ergonomie du questionnaire CAWI. En effet plusieurs aspects rendaient la passation laborieuse.

- une question par page
- un bouton validé peu accessible : souvent utilisation de la barre de défilement pour atteindre, le bouton validé en bas de page.
- Des boutons radios difficiles à cliquer
- Lenteur du système
- Peu esthétique
- pas de titre indiquant dans quelle partie du questionnaire on se situe, ce qui complique la compréhension du déroulement du questionnaire.

Ensuite des abandons beaucoup plus massifs pour certaines questions, trois cas différents ont été repérés concernant les abandons massifs.

- Les questions avec consignes, qui augmentent la charge pour le répondant et par conséquent le risque d'abandon.
- pour la saisie du calendrier professionnel on demande que les séquences consécutives soient de natures différentes, dans le cas contraire on demande à l'enquêté de corriger. Hors la correction de la séquence exige de ressaisir l'intégralité de l'information, (date de début, date de fin, situation, commune). Cette demande de correction occasionne de nombreux abandons.
- Beaucoup d'abandon à la fin de la description de la première séquence du calendrier. Ces abandons concernent exclusivement les sortants dont la première séquence est inférieure à 4 mois. Il pourrait s'agir des sortants dont le parcours à l'issue du système éducatif comporte un grand nombre de séquences différentes, de courte durée, et qui jugent trop lourde la description de l'intégralité de leur parcours.

2 Première approche standard : correction de la non réponse et calage

Du fait des défauts de couverture et des abandons en cours lié au moins au niveau de formation, il apparaît que les données de la population CATI et de la population CAWI ne sont pas directement comparables. En effet, du fait de l'absence de mail dans les bases du rectorat, plus de 90% des répondants CAWI sont des sortants de l'enseignement supérieur. D'autre part, au sein du supérieur, certaines strates de formations fines contiennent très peu voir pas de répondant (toujours du fait de l'absence de mail dans les bases des établissements correspondants). Il a donc été décidé de restreindre la population aux strates de formation qui présentent suffisamment d'effectif, c'est-à-dire un sous ensemble des strates de formation du supérieur. La correction de la non-réponse et le calage sont faits sur cette sous population.

Du fait de la présence de nombreux hors champ, la correction de la non-réponse est décomposée en plusieurs étapes imbriquées. La première concerne la probabilité que le sortant reçoive un email, la seconde que le sortant finisse le questionnaire filtre c'est-à-dire qu'il soit identifié comme hors champ ou qu'il soit dans le champ et la dernière concerne la modélisation de la probabilité de compléter le questionnaire. Les modèles logistiques mobilisent uniquement les

variables présentes dans la base de sondage qui sont en nombre limité. Avec des aires sous la courbe de ROC de l'ordre de 0.60, les modèles ne sont donc que moyennement explicatif.

A l'issue de la correction de la non-réponse, des différences importantes subsistent entre les structures des répondants par CATI et des répondants par CAWI. Pour corriger ces différences un calage sur marge est effectué. La démarche adoptée est de faire un calage du CAWI sur la structure du CATI par rapport aux variables qui expliquent le mieux les variables essentielles de l'enquête.

3 Comparaison entre les réponses internet et téléphone

La cohérence entre les structures des réponses internet et téléphone n'est que partielle et différents cas de figure apparaissent : les variables qui présentent des différences de structures importantes du fait de la différence de passation. En effet pour certaines questions la modalité « Ne sais pas » est proposée d'emblée pour la passation par internet alors qu'elle n'est pas énoncée lors de la passation par téléphone. Pour les questions sensibles ou d'opinion, ces différences de présentation de modalités peuvent occasionner un phénomène de désirabilité sociale : les répondants par téléphone qui n'ont a priori pas la possibilité de ne pas se prononcer semblent privilégier la réponse la plus convenue. Ensuite d'autres questions qui ne sont pas affectés a priori par des écarts d'énonciation présentent tout de même des structures différentes, qui peuvent être dues aux effets de sélection non corrigés ou aux différences de perception de l'information entre les passations internet et téléphone.

4 Calcul de précision des estimations

Pour affirmer la significativité de ces différences il est nécessaire d'estimer la variance des statistiques produites afin de ne pas interpréter en tant qu'effets de mesure ou effets de sélection des écarts principalement dus à la variance des estimations.

Cette estimation de la variance a été faite par calcul direct. En collaboration avec la direction de la méthodologie de l'Insee, nous estimons la variance due à la composition des différentes étapes mise en œuvre.

Nous allons procéder en trois étapes :

- Une première estimation de variance a été faite à partir de l'approximation de variance pour les tirages équilibrés de Deville et Tillé
- La prise en compte de l'ajout de variance due à la non réponse a été faite en considérant le mécanisme de non réponse comme un tirage à deux degrés, le deuxième degré correspondant ici au phénomène de non-réponse. Le terme additionnel est déterminé en utilisant la formule de Rao.
- La prise en compte de l'effet du calage sur la variance de l'estimateur se fait en considérant que l'estimateur calé a les mêmes propriétés que l'estimateur par la régression. De ce fait ce sont les variances des résidus des variables d'intérêt après régression sur les variables de calage qui sont pris en compte dans le calcul de variance.

Étant donné que nous travaillons essentiellement sur des variables catégorielles ce sont les variances des proportions qui sont estimées à partir des variables linéarisées.

5 Mobilisation d'une méthode de matching

Pour essayer de prendre en compte les effets de sélection des effets de mesure il est possible d'appliquer à la méthodologie d'enquête des méthodes économétriques utilisées pour évaluer les effets des dispositifs dans le cadre de politiques publiques. Idéalement pour mesurer l'effet d'une politique publique il faut qu'il y ait tout d'abord une association aléatoire entre les prétendants et le dispositif pour éviter de confondre les effets du dispositif avec des caractéristiques individuelles prises en compte lors de la sélection. Lorsqu'un dispositif est mis

en place, sans cette expérimentation aléatoire préalable, les méthodes de matching sont utilisées pour théoriquement contrôler les effets de sélection et déterminer si les bénéficiaires du dispositif voient leur situation changer par rapport à la variable d'intérêt. Dans le cadre de la comparaison des modes de collectes, le dispositif à évaluer est la passation par CAWI.

La variable de traitement T est égale à 1 pour les enquêtés répondant par CAWI et 0 par téléphone, les répondants par téléphone constituent ici la population de contrôle. La variable d'intérêt est notée Y . Deux variables latentes sont associées à cette variable d'intérêt, $Y1$ et $Y0$ qui correspondent aux valeurs potentiellement renseignées par le répondant respectivement par internet et par téléphone. Ces variables latentes ne sont observées que partiellement et on a la relation :

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

L'hypothèse de cette approche par matching est de considérer que conditionnellement aux variables qui expliquent la sélection les variables latentes sont indépendantes du traitement. D'après la propriété de Rubin et Rosenbaum, les individus seront appariés par leur score de propension. Les écarts moyen entre les valeurs des individus CAWI et celles de leurs contrefactuels mobiliseront des estimateurs à noyau.

La méthode de matching permet de mobiliser en plus des variables initiales de la base de sondage, les réponses fournies par les enquêtés. On fait l'hypothèse que les variables explicatives issues des réponses ne sont pas sensibles au mode, c'est-à-dire que les sortants répondent de la même manière quel que soit le mode. Cette méthode sera appliquée sur les variables les plus centrales du questionnaire et celles montrant le plus d'écart entre les deux modes de collectes.

Que ce soit pour cette partie ou la précédente concernant la correction de la non-réponse et le calage, il est difficile de conclure sur la nature des écarts observés entre les deux collectes et de dissocier les effets de sélection des effets de mesure.

6 Conclusion

Malgré les difficultés rencontrées lors de la collecte (difficulté à enrichir la base de sondage, mauvaise ergonomie du questionnaire web, ...), les premières comparaisons sur des indicateurs centraux du dispositif (fortement lié au niveau de diplôme notamment) sont plutôt encourageants dès lors qu'une correction sur la sélection est mise en œuvre. Passé ce premier constat, des différences significatives subsistent la majorité des variables.

Le calcul de précision des estimations permet d'identifier les variables pour lesquelles les différences sont significatives. Il s'agira alors de réfléchir à la formulation des questions pour véhiculer la même information par CAWI et par CATI. Une démarche intéressante, mais coûteuse, pour valider des formulations alternatives et de fournir différentes version du questionnaire, et d'utiliser des méthodes SEM pour estimer l'impact de la manière de poser la question sur la structure de la réponse.

Pour étudier les effets de sélection des effets de mesure, la mise en place d'un échantillon expérimental est souvent préconisée. Les individus appartenant à cet échantillon et acceptant de répondre seront associé aléatoirement à l'un des deux modes de collectes afin de supprimer les effets de sélection.

Bibliographie

- [1] Anna-Maria Schielicke. (s.d.). "Don't know" the difference An experimental comparison between Web and CATI.
- [2] Annette, J., Caroline, R., & Peter, L. (2008, February). Assessing the Effect of Data Collection Mode on Measurement.
- [3] Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., et al. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet.
- [4] Frippiat, D., & Marquis, N. (2010). les enquêtes par internet en sciences sociales: un état des lieux. population.
- [5] Huguée, C., & Kern, A.-L. (s.d.). Observer les télénquêteurs. Les paradoxes de la rationalisation de la production statistique. Genèses, 2008/3 n° 72, , pp. p. 102-118.
- [6] Krosnick, & jon, L. C. (2010). Comparing oral interviewing with self-administrated computerized questionnaires an experiment.
- [7] Leeuw, E. D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. Journal of Official Statistics, pp. 233-255.
- [8] Leeuw, E. D., & Hox, j. (s.d.). Internet Surveys as Part of a Mixed-Mode Design.
- [9] Perreti, G. d. (s.d.). Elfe, le multimode "fantasy". Séminaire de méthodologies statistiques: les enquêtes multimode.
- [10] Vannieuwenhuyze, & Jorre. (2010). A METHOD FOR EVALUATING MODE EFFECTS IN MIXED MODE SURVEY. Public opinion quaterly.
- [11] Revilla, Saris, (2010). A comparison of the quality of ESS questions in different data collection modes