

ENQUETE 2010 AUPRES DE LA GENERATION 2007 : TAUX D'INSISTANCE ET CARACTERISTIQUES DES REpondANTS

Pascale ROUAUD¹ ()*

() Céreq, Département des Entrées et Evolutions dans la Vie Active*

Introduction

Les enquêtes « Génération » du Centre d'études et de recherches sur les qualifications (Céreq) permettent d'étudier les conditions d'accès à l'emploi des jeunes sortants du système éducatif durant leurs premières années de vie active en fonction de leur formation initiale. Tous les trois ans, une nouvelle cohorte de jeunes sortis du système éducatif au cours ou à la fin d'une année scolaire donnée est interrogée trois ans après sa sortie sur les parcours depuis la sortie. Certaines cohortes sont réinterrogées ultérieurement.

La base de sondage est constituée à partir de listes d'inscrits fournies par les établissements de formation. Elle contenait environ 1 200 000 individus pour l'enquête 2010, dont une partie n'entrait pas dans le champ de l'enquête parce qu'ils avaient poursuivi leurs études après l'année scolaire considérée, sans qu'il soit possible de le savoir a priori. Le questionnaire s'est déroulé sous forme d'un entretien téléphonique, les coordonnées téléphoniques des jeunes échantillonnés étant recueillies auprès de leur établissement de formation puis enrichies par des recherches à partir des fichiers de France Telecom. Cette phase a conduit à recueillir aussi des numéros de proches ou d'homonymes. On a finalement abouti à un total de 1 à 9 numéros de téléphone pour chaque jeune envoyé en plateau de collecte.

Environ 740 000 jeunes sont sortis du système éducatif au cours ou à la fin de l'année scolaire 2006-2007. L'enquête 2010 conduite auprès de la « Génération 2007 » portait sur ces jeunes. L'échantillon utilisé comprenait 192 000 jeunes, dont une partie hors du champ de l'enquête (surtout des individus ayant poursuivi leurs études après 2007, pour lesquels cette information était recueillie lors de la passation du questionnaire filtre introductif). A l'issue de la collecte, environ 24 500 questionnaires avaient été réalisés dans le champ de l'enquête du Céreq.

Notre objectif est d'étudier les liens entre les caractéristiques des coordonnées téléphoniques (fixe, mobile, origine du numéro), la gestion des appels et relances téléphoniques par le prestataire sélectionné pour réaliser la collecte et le résultat final des appels par individu. Les caractéristiques des répondants au regard de ces différents paramètres seront analysées afin de mettre en évidence d'éventuels biais sur les variables principales observées sur les répondants (sexe, niveau de formation, situation d'activité). Il s'agit ici pour le producteur de vérifier l'utilité de la phase d'enrichissements en numéros de téléphone et de détecter les biais dans les résultats liés à la gestion des appels.

¹ Je tiens à remercier Bastien Thome, pour les pistes de réflexions ouvertes par son rapport de stage. Je remercie également Stéphane Jugnot, Jean-Claude Sigot et Philippe Winnicki, pour leurs remarques et conseils sur cette étude.

Le papier est organisé de la manière suivante. La première partie est consacrée à la présentation des données disponibles dans l'échantillon utilisé, au déroulement et aux résultats généraux de la collecte. La seconde partie s'intéresse à la manière dont le prestataire a exploité les coordonnées téléphoniques disponibles ainsi qu'aux résultats d'appels. Dans la troisième partie, nous étudions les relations entre les caractéristiques d'insertion et le rang d'appel ayant conduit à la réalisation du questionnaire.

1. Présentation des données et résultats généraux

1.1. Informations présentes dans la base de sondage et l'échantillon

La base de sondage de « Génération 2007 » a été constituée à partir des informations nominatives transmises par les établissements de formation sur les sortants potentiels de formation initiale. Pour chaque individu scolarisé pendant l'année scolaire 2006-2007 et présumé sortant du système éducatif, les informations suivantes ont été collectées : nom et prénom, date de naissance, sexe, diplôme et spécialité de formation, adresse postale de 2007. Pour 75% des individus présents dans la base de sondage, un ou plusieurs numéros de téléphone ont également été renseignés par les établissements de formation.

L'échantillon utilisé contenait 192 000 individus, tirés aléatoirement au sein de la base de sondage, afin de satisfaire les besoins du Céreq et de ses partenaires. Parmi eux, 113 200 (59%) disposaient d'au moins un numéro de téléphone issu de la base de sondage. Un ou plusieurs numéros de téléphone portable étaient renseignés pour 67 500 individus.

Le Céreq a enrichi les informations issues de la base de sondage. Il a ainsi effectué une mise à jour des coordonnées postales, à la date de l'enquête, afin d'obtenir la nouvelle adresse des individus ayant déménagé entre 2007 et 2010. Une recherche de coordonnées téléphoniques a ensuite été réalisée à partir des fichiers de France Télécom². La première phase de recherche (phase A) consistait à trouver un écho unique sur les nom, prénom et adresse de l'individu. Cette phase a permis de trouver 57 200 individus (30%), dont 31 200 ne disposaient pas de numéro de téléphone dans la base de sondage.

En cas d'échec de la phase A, des recherches de coordonnées élargies étaient mises en œuvre (un ou plusieurs échos³ pour retrouver l'individu sur les nom-prénom-commune, nom-prénom-département, nom-prénom-région, nom-prénom-Ile de France ou pour retrouver un membre de la famille sur le nom-commune).

L'ensemble des numéros de téléphone étaient ensuite hiérarchisés selon leur fiabilité, afin que le prestataire d'enquête appelle en priorité les numéros les plus fiables (voir infra).

Parmi les 192 000 individus échantillonnés, 167 700 (87%) disposaient d'un ou plusieurs numéros de téléphone à l'issue de l'ensemble des recherches de coordonnées téléphoniques. Un total de 24 300 individus (13%) a été définitivement perdu car le Céreq ne disposait d'aucune coordonnée téléphonique à l'issue de l'ensemble de ses recherches.

² L'absence d'annuaire universel constitue un véritable handicap pour les enquêtes téléphoniques telles que Génération. En effet, la proportion des individus non répertoriés dans les fichiers de France Télécom est de plus en plus importante (individus possédant uniquement un téléphone portable ou clients chez d'autres opérateurs) et il n'existe pas de moyen standardisé d'obtenir leurs coordonnées.

³ Le Céreq a limité ses recherches à un total de cinq échos par individu sur les recherches de coordonnées élargies.

Tableau 1. Types de coordonnées téléphoniques

	Fréquence	Pourcentage
Uniquement base de sondage	41 200	21%
Uniquement recherche phase A	31 200	16%
Uniquement recherche élargie	22 700	12%
Base de sondage et phase A	26 000	14%
Base de sondage et recherche élargie	46 000	24%
Recherche manuelle Céreq ⁴	600	0%
Aucune coordonnée disponible	24 300	13%
Total	192 000	100%

Source : enquête 2010 auprès de la Génération 2007

Champ : échantillon utilisé

Les individus disposant d'au moins un numéro de téléphone en avaient en moyenne 2,2 (min : 1 – max : 8, ET= 1,43) dont 0,9 en moyenne issu des recherches téléphoniques élargies (min : 0 – max : 5, ET=1,52). 97 800 individus disposaient d'au moins deux numéros de téléphone.

Tableau 2. Nombre de numéros disponibles par individu

Nombre de numéros disponibles par individu	Fréquence	Pourcentage
Aucun	24 300	13
1	69 900	36
2	48 200	25
3	20 600	11
4 et plus	29 000	15
Total	192000	100

Source : enquête 2010 auprès de la Génération 2007

Champ : échantillon utilisé

1.2. Déroulement de la collecte

La collecte de l'enquête s'est déroulée par téléphone du 6 avril au 10 juillet 2010. Le Céreq a transmis au prestataire en charge de cette collecte le fichier relatif aux 167 700 individus disposant d'un ou plusieurs numéros de téléphone.

La majorité des individus échantillonnés ont été envoyés directement en plateau d'enquête, afin de réaliser immédiatement les interrogations téléphoniques. Cependant, un terrain de collecte spécifique appelé terrain de pré-qualification, a été mis en place pour les individus issus de formations connues pour leur fort taux de hors champ (dont la probabilité d'appartenir au champ de l'enquête était estimée inférieure à 15%) ou pour les individus dont les coordonnées téléphoniques étaient jugées peu fiables. Un total de 29 000 individus échantillonnés ont été envoyés en pré-qualification. Cette phase de premier contact permettait de valider l'appartenance au champ et/ou les coordonnées téléphoniques. Seuls les individus pour lesquels ces informations étaient validées étaient par la suite envoyés en plateau d'enquête.

⁴ En cours d'enquête, le taux de réponse des sortants de formations du sport et de l'animation s'est avéré inférieur aux prévisions. Une recherche manuelle sur les sortants de ces formations pour lesquels on ne disposait d'aucune coordonnée téléphonique à l'issue des recherches élargies a été effectuée, permettant de trouver des coordonnées pour plus d'un tiers d'entre eux.

Pour les individus disposant de plusieurs numéros de téléphone, les numéros ont été hiérarchisés par ordre de fiabilité décroissante. Les numéros considérés comme les plus fiables étaient ceux issus de la base de sondage, puis ceux issus des recherches de coordonnées téléphoniques phase A et enfin les numéros de téléphones issus des phases de recherches élargies. Lorsque l'établissement de formation avait transmis un numéro de portable et un numéro de fixe pour le même individu, le numéro portable était prioritaire.

En début d'enquête, le numéro le plus fiable était envoyé en collecte et les numéros suivants étaient temporairement bloqués. Lorsqu'un appel permettait le classement de l'individu (réalisation de l'enquête, mise hors champ, indisponibilité durant toute la période de collecte ou refus), les tentatives étaient stoppées sur l'ensemble des numéros disponibles pour cet individu. Lorsqu'un appel permettait de classer le numéro de téléphone mais pas l'individu (contact avec une personne qui n'était pas l'individu recherché), les tentatives étaient stoppées sur le numéro en exploitation et le numéro suivant, lorsqu'il existait, devait être tenté.

Le prestataire avait pour consigne d'appliquer un coefficient d'insistance de dix tentatives d'appels par numéro de téléphone. Si les dix premières tentatives sur le numéro de priorité 1 se révélaient infructueuses, ce numéro était abandonné. Le cas échéant, le prestataire passait au numéro suivant et ainsi de suite jusqu'à épuiser l'ensemble des numéros de téléphone disponibles.

Les règles de rappel étaient gérées par un automate d'appels, afin de maximiser les chances de joindre l'individu (tentatives sur des plages horaires et des jours de la semaine différents,...) sur un numéro donné.

1.3. Résultats globaux

A l'issue du terrain d'enquête, chaque individu échantillonné a été classé de la manière suivante :

- **Succès** lorsque le Céreq a réussi à joindre l'individu recherché et que ce dernier a accepté de participer à l'enquête. Il s'agit des individus qui ont répondu à l'enquête ainsi que des individus qui se sont révélés hors champ à l'issue du questionnaire filtre (en poursuite d'études après 2007 ou non primo-sortants).
- **Échec** lorsque le Céreq ne disposait d'aucune coordonnée téléphonique, qu'il était impossible d'enquêter l'individu recherché (refus, injoignable durant la période de collecte, personne résidant à l'étranger,...) ou que l'ensemble des numéros de téléphone avait été invalidé (faux numéro, mauvais interlocuteur, refus de tiers,...)
- **Non classé** dans tous les autres cas, c'est-à-dire pour les individus échantillonnés disposant d'un ou plusieurs numéros de téléphone, non répondants à l'issue du plateau de collecte et pour lesquels des tentatives d'appels téléphoniques restaient envisageables⁵.

Le taux de succès, correspondant au taux réponse de l'enquête 2010 auprès de la « Génération 2007 », est de 25%. Sont donc considérés comme répondants les individus que le Céreq a réussi à contacter et qui ont accepté de répondre, c'est-à-dire les individus ayant effectivement répondu à l'ensemble du questionnaire ainsi que les individus classés hors du champ de l'enquête à l'issue du questionnaire filtre. La proportion d'individus hors du champ était élevée et le taux de réalisation d'enquête dans le champ faible puisque seuls 13% des individus échantillonnés ont donné lieu à la réalisation d'une enquête complète.

Le taux d'échec a atteint 30%. Il s'explique principalement par la mauvaise qualité des coordonnées disponibles (aucune coordonnée ou coordonnée erronée). Les taux de refus direct ou refus d'une tierce personne étaient faibles et concernaient 5% de l'échantillon utilisé.

Enfin, la proportion d'individus non classés à l'issue de la collecte a atteint 45%.

⁵ Les individus pour lesquels plus de 10 tentatives infructueuses ont été réalisées sur le même téléphone restent dans la catégorie Non classé. En effet, même si le nombre d'appels préconisé a été réalisé, d'autres tentatives restaient envisageables dans l'absolu.

Tableau 3. Détail du classement des individus échantillonnés

Classement	Fréquence	Pourcentage
Succès	48 500	25%
Enquête réalisée dans le champ de l'enquête Céreq	24 600	13%
Individu hors du champ de l'enquête Céreq	23 900	12%
Echec	58 000	30%
Aucune coordonnée disponible	24 300	13%
Refus de tiers, tous numéros tentés	3 000	2%
Refus de l'individu	5 900	3%
A déjà répondu	400	0%
Décédé	400	0%
Réside à l'étranger	1 200	1%
Injoignable pendant la période d'enquête	3 000	4%
Mauvais interlocuteur, tous numéros tentés	15 500	8%
Interlocuteur qui ne connaît pas les coordonnées de la personne recherchée, tous numéros tentés	3 800	2%
Pas d'interlocuteur : faux numéro, disque France Telecom, tous numéros tentés	500	0%
Non classé à l'issue du plateau d'enquête	85 500	45%
Pas d'interlocuteur : ne répond pas, occupé,...	46 900	24%
A rappeler, absent	24 700	13%
Abandon en cours d'enquête	1 800	1%
Refus de tiers, numéros encore disponibles	700	0%
Mauvais interlocuteur, numéros encore disponibles	5 800	3%
Interlocuteur qui ne connaît pas les coordonnées de la personne recherchée, numéros encore disponibles	900	0%
Pas d'interlocuteur : faux numéro, disque France Telecom, numéros encore disponibles	1 800	1%
Individu non appelé	2 900	2%

Source : enquête 2010 auprès de la Génération 2007

Champ : échantillon utilisé

2. Exploitation de l'échantillon et résultats, statistiques descriptives

L'analyse présentée par la suite concerne uniquement sur les 137 500 individus directement envoyés en plateau d'enquête⁶. Ce choix a été motivé par la volonté de travailler sur des données homogènes. Les individus passés par le terrain de pré-qualification sont donc exclus de l'analyse. D'une part, les individus passés par la phase de pré-qualification présentent des caractéristiques très spécifiques, d'autre part, leur passage en plateau d'enquête est conditionné par le résultat de la phase de pré-qualification. Pour eux, une analyse spécifique et plus complexe serait donc nécessaire.

2.1. Types de coordonnées téléphoniques disponibles

Un total de 287 600 numéros de téléphone était disponible pour les individus directement envoyés en enquête. Parmi eux, 151 800, soit 53% ont finalement été exploités. Les numéros de téléphone portable issus des établissements de formation ont presque toujours été utilisés. Ce type de coordonnée téléphonique, considéré comme le plus fiable, a été placé en ordre de priorité 1. Ensuite, les numéros de téléphone fixe issus des établissements de formation et les numéros issus des recherches téléphoniques phase A ont été largement utilisés. Les numéros issus des recherches téléphoniques élargies, considérés comme les moins fiables ont été utilisés dans moins de 10% des cas.

Tableau 4. Origine des coordonnées téléphoniques disponibles

Origine du numéro de téléphone	Nombre de numéros de téléphone	Nombre d'individus concernés	Taux d'utilisation
Etablissement de formation : téléphone fixe	63 200	61 400	63%
Etablissement de formation : téléphone portable	66 800	65 200	99%
Recherche phase A, hors individu ayant déménagé	48 600	48 600	71%
Recherche phase A sur individu ayant déménagé	2 800	2 800	57%
Recherche élargie : Nom - prénom - commune	300	100	0%
Recherche élargie : Nom - prénom - département	6 300	4 500	9%
Recherche élargie : Nom - prénom - région	27 800	16 700	8%
Recherche élargie : Nom - commune	27 500	16 000	10%
Recherche élargie : Nom - prénom - Ile de France	8 400	5 300	8%
Recherche élargie : Nom - département	35 300	17 300	8%
Recherche manuelle effectuée par le Céreq	600	600	100%
Total	287 600		53%

Source : enquête 2010 auprès de la Génération 2007

Champ : échantillon des 137 500 individus directement envoyés en enquête

Lecture : parmi les numéros de téléphone disponibles, 63 200 étaient des numéros fixes transmis par les établissements de formation. 61 400 individus possédaient au moins un numéro fixe transmis par son établissement de formation. 63% des numéros de fixe ont été effectivement utilisés par le prestataire pendant la collecte (au moins une tentative d'appel).

Pour 47% des individus directement envoyés en enquête, les établissements de formation avaient fourni un numéro de téléphone portable. La présence d'un numéro de portable variait fortement selon le type de formation suivie : les portables étaient plus présents dans les formations de l'enseignement supérieur et très rares pour les collèges et lycées. On disposait d'un numéro de portable pour 52% des filles et 42% des garçons. Pour un type de formation donné, les filles disposaient plus souvent d'un numéro de portable que les garçons.

⁶ Sont considérés ici uniquement les individus pour lesquels le prestataire a réalisé au moins une tentative d'appel.

Tableau 5. Présence de numéros de portable selon la formation suivie

Type de formation suivie en 2007	Nombre d'individus concernés	Pourcentage de ceux pour lesquels on dispose d'un portable
Collège	1900	3%
CAP ou BEP voie scolaire	8300	11%
CAP, BEP, lycée par apprentissage	17400	46%
Lycée professionnel ou technologique	11900	5%
Bac+1 à bac+5 par apprentissage	4500	58%
BTS ou DUT	12100	28%
Autres formations Universitaires (bac+1 à bac+8)	45900	62%
Formations de la jeunesse et des sports	7400	31%
Formations de la santé et du social	19700	68%
IUFM	2200	86%
Ecoles de l'enseignement supérieur (dont commerce et ingénieurs)	6200	61%
Total	137500	53%

Source : enquête 2010 auprès de la Génération 2007

Champ : échantillon des 137 500 individus directement envoyés en enquête

Pour 123 400 individus directement envoyés en enquête (89,8%), seul un numéro de téléphone a finalement été exploité. Deux numéros ont été exploités pour 13 700 individus (10%) et trois ou quatre numéros pour 400 individus (0,2%).

Par la suite, nous rattachons chaque numéro de téléphone utilisé à son origine. Par exemple, imaginons que le prestataire ait tenté un premier appel sur un téléphone fixe transmis par l'établissement de formation. Il tombe sur le père de l'individu recherché qui communique un numéro de téléphone portable pour joindre son fils. Les appels ultérieurs sur le portable seront rattachés au numéro d'origine, c'est-à-dire le fixe transmis par l'établissement de formation. En effet, c'est le point d'entrée qui permettra (ou non) d'aboutir à un succès qui nous intéresse.

2.2. Exploitation de l'échantillon

2.2.1. Première tentative d'appel sur le numéro prioritaire

La collecte a duré 15 semaines (du 6 avril au 10 juillet 2010). L'exploitation de l'échantillon s'est déroulée de manière progressive. Ainsi, après trois semaines d'enquête, le prestataire avait effectué une tentative d'appel auprès de seulement 25% des individus directement envoyés en enquête. Après huit semaines d'enquête, au moins un appel avait été tenté auprès de l'ensemble des individus appartenant à l'échantillon principal, soit 76% des individus échantillonnés directement envoyés en enquête. Les premiers appels sur le numéro de priorité 1 effectués entre la semaine 9 et 15 correspondent à l'injection, en cours de collecte, de l'échantillon de réserve⁷.

Graphique 1. Calendrier des premiers appels sur le numéro de priorité 1 pour l'échantillon des individus directement envoyés en enquête

Semaine	Échantillon principal (%)	Échantillon de réserve (%)	Total (%)
1	0	0	0
2	15	0	15
3	25	0	25
4	35	0	35
5	45	0	45
6	55	0	55
7	65	0	65
8	76	0	76
9	76	0	76
10	76	0	76
11	76	4	80
12	76	14	90
13	76	14	90
14	76	24	100
15	76	24	100

Source : enquête 2010 auprès de la Génération 2007

Champ : échantillon des 137 500 individus directement envoyés en enquête

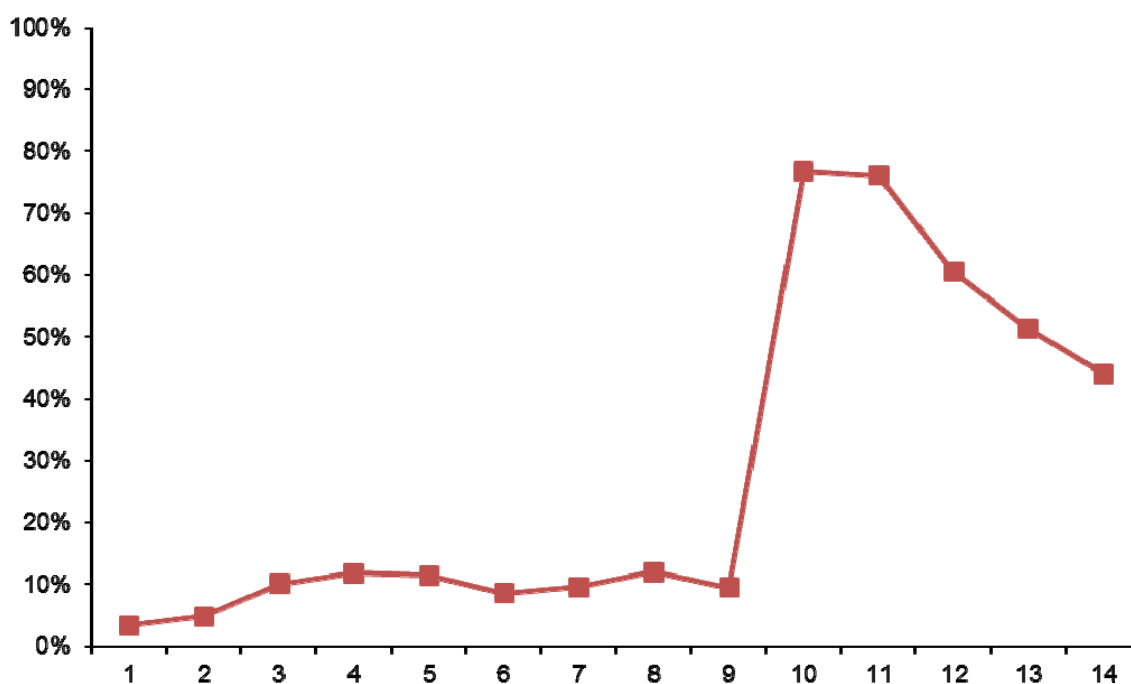
⁷ Avant le début de la collecte et afin de pouvoir faire face à un taux de réponse inférieur aux attentes, le Céreq avait constitué un échantillon de réserve. Une utilisation partielle de cet échantillon a été effectuée en cours de collecte sur des cibles particulières.

2.2.2. Persévérance du prestataire

Un total de 681 800 tentatives d'appels a été réalisé par le prestataire, sur les 151 800 numéros de téléphone exploités. Seuls 4 000 numéros de téléphone (3%) ont fait l'objet d'au moins 11 tentatives d'appels.

La consigne était donnée au prestataire d'appliquer un coefficient d'insistance de dix tentatives par numéro de téléphone, en cas d'appel infructueux, avant d'abandonner ce numéro et de passer, le cas échéant, au suivant. On constate qu'en pratique, la consigne n'a souvent pas été respectée et que le taux d'abandon avant la dixième tentative est élevé. En moyenne, le prestataire a ainsi réalisé seulement 6,3 tentatives infructueuses avant d'abandonner un numéro. Un total de 44 700 numéros de téléphone a donc été abandonné avant d'avoir réalisé les dix tentatives d'appels préconisées.

Graphique 2. Taux d'abandon en fonction du rang de l'appel



Source : enquête 2010 auprès de la Génération 2007

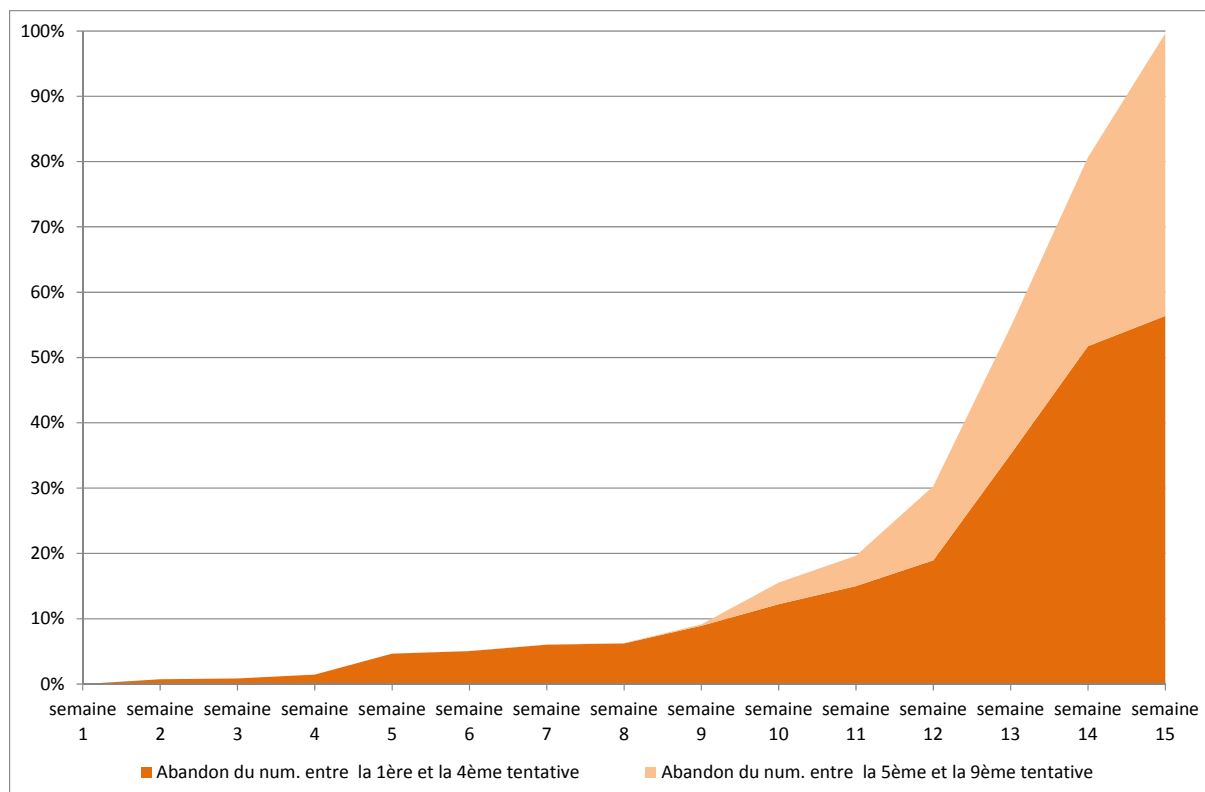
Champ : ensemble des appels infructueux effectués auprès des 137 500 individus directement envoyés en enquête

Lecture : après la 3^{ème} tentative d'appel infructueuse sur un numéro, le prestataire a abandonné dans 10% des cas.

Le taux d'abandons après chaque tentative infructueuse est inférieur à 5% pour les deux premières tentatives, oscille autour de 10% pour les tentatives trois à neuf et connaît, de façon logique, son pic à l'issue de la dixième tentative d'appel. A ce moment, 77% des numéros sont abandonnés par le prestataire.

Pour 90% des numéros abandonnés avant d'avoir atteint les dix tentatives, le résultat du dernier appel est ne répond pas, répondeur ou occupé. Les numéros de téléphone ont essentiellement été abandonnés en fin de collecte : 10% l'ont été lors des deux premiers mois de collecte et 45% lors des trois dernières semaines. Le graphique 3 suggère une exploitation insuffisante du fichier par le prestataire durant la collecte, se concluant en fin de période par la non réalisation du nombre de tentatives préconisées sur de nombreux numéros.

Graphique 3. Calendrier des abandons avant la dixième tentative d'appel



Source : enquête 2010 auprès de la Génération 2007

Champ : ensemble des appels infructueux effectués auprès des 137 500 individus directement envoyés en enquête

Lecture : parmi les numéros abandonnés, 20% l'avaient été avant la fin de la 11^{ème} semaine de collecte (15% entre la 1^{ère} et la 4^{ème} tentative et 5% entre la 5^{ème} et la 9^{ème} tentative).

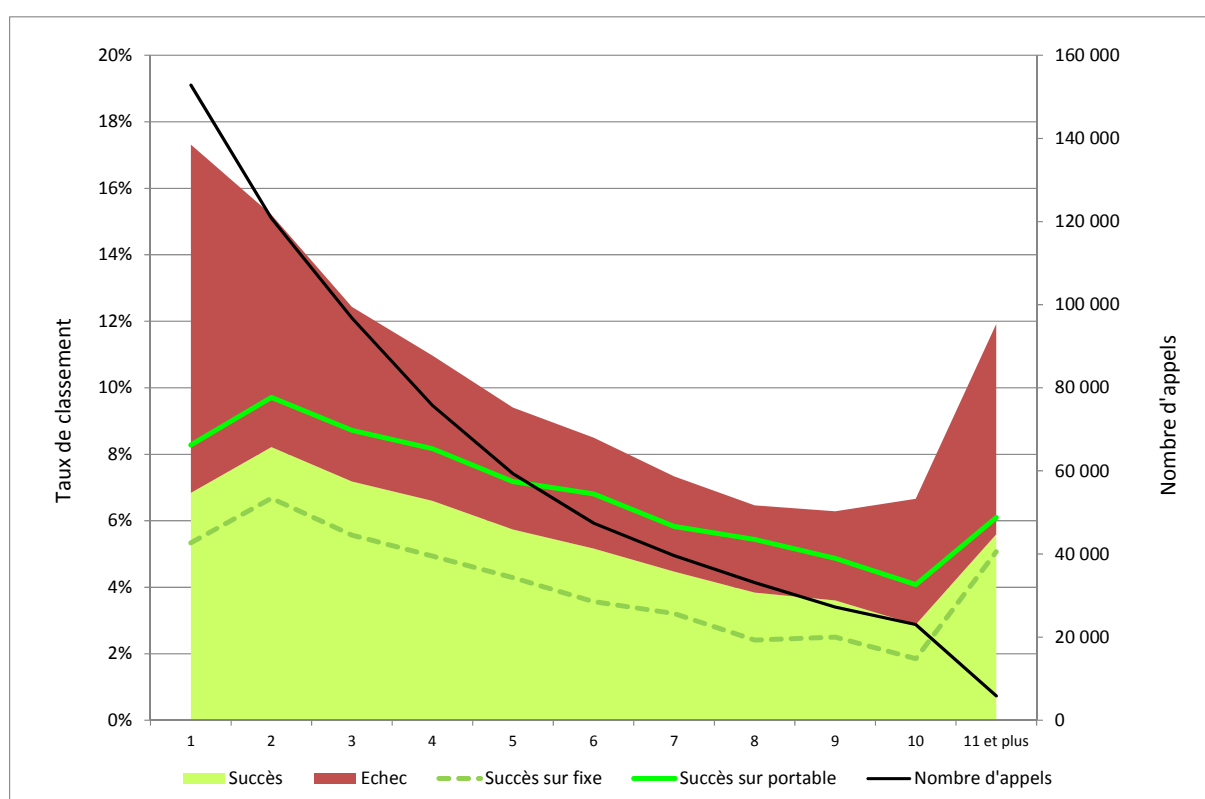
2.2.1. Résultat en fonction des caractéristiques de l'appel

On étudie maintenant les résultats de l'ensemble des 681 800 appels émis par le prestataire. Tous les numéros pour lesquels au moins une tentative a été réalisée sont concernés, ainsi que tous les rangs d'appel.

Le taux de classement moyen par tentative d'appel était de 12,3%. Chaque appel avait ainsi 6,3% de chances de se conclure par un succès en moyenne et 6% de chances d'aboutir sur un échec. Les appels effectués sur les téléphones portables étaient nettement plus efficaces que ceux effectués sur les téléphones fixes, 7,8% des appels passés sur un portable se terminant par un succès (4,8% pour les fixes) et 5,9% par un échec (6,9% pour les fixes).

Dans 1,4% des cas, les tentatives d'appel se sont conclues par un échec suite à un refus direct ou d'une tierce personne. Ce type de refus était plus fréquent pour les numéros de téléphone fixe (1,8% des tentatives d'appels) que pour les portables (0,9%).

Graphique 4. Classement des appels en fonction du rang de l'appel



Source : enquête 2010 auprès de la Génération 2007

Champ : ensemble des tentatives d'appels effectuées auprès des 137 500 individus directement envoyés en enquête

Lecture : une première tentative d'appel a été effectuée sur 151 800 numéros de téléphone. Elle a permis le classement de 17,3% des numéros (6,8% en succès et 10,5% en échec). Le taux de succès était de 8,3% pour les portables et 5,3% pour les fixes.

Le rendement des appels effectués sur un même numéro de téléphone a été décroissant jusqu'à la dixième tentative. Cependant, on n'observe pas de décrochage indiquant qu'à partir d'un certain nombre de tentatives, les appels deviendraient inutiles. Ainsi, le taux moyen de succès à la dixième tentative était de 2,9%. A chaque rang de tentative, les téléphones portables ont permis d'obtenir de meilleurs résultats que les fixes. A la première tentative, les portables donnaient lieu à 8,3% de succès contre 5,3% pour les fixes. A la dixième tentative d'appel, ces résultats tombaient respectivement à 4,1% et 1,9%. Au-delà de la dixième tentative, le taux de classement a augmenté. En effet, le prestataire a consenti à réaliser des tentatives supplémentaires sur des individus spécifiques tels que

ceux qui avaient pris un rendez-vous pour réaliser l'enquête lors de la dixième tentative ou avaient demandé à être rappelés ultérieurement.

Le prestataire a effectué des appels du lundi au samedi, entre 10 heures et 20h30 (17 heures le samedi). Le jour de la semaine où le prestataire a réalisé le maximum de tentatives d'appels et a classé le plus grand nombre d'individus en succès est le samedi. Concernant les plages horaires, la tranche horaire comprise entre 17 et 19 heures a été la plus exploitée et a permis d'enregistrer 50% des succès globaux.

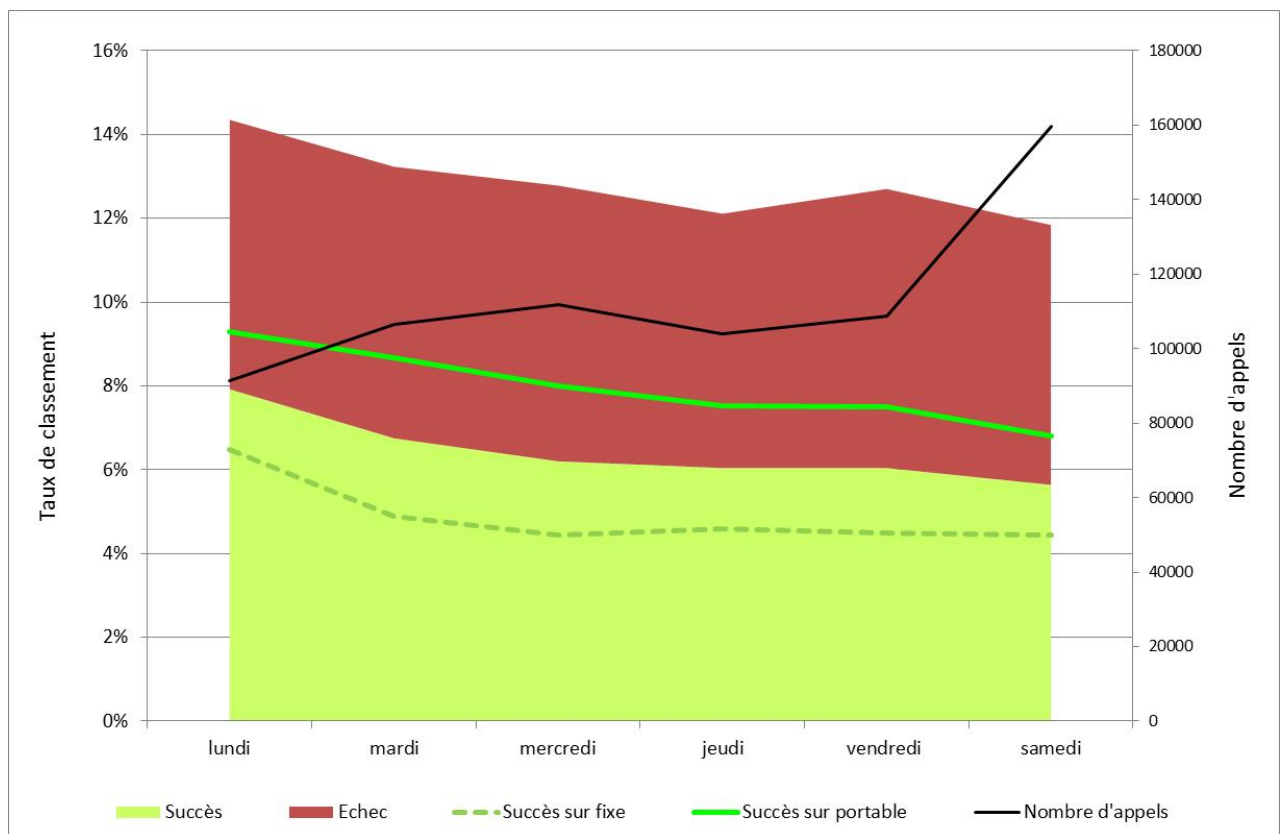
On constate que le taux de classement varie selon le jour et l'horaire de la tentative d'appel. On n'observe cependant pas de corrélation positive entre le taux de classement et le volume d'appels émis.

Ainsi, la journée du lundi, qui restait la moins exploitée par le prestataire, est pourtant celle qui présentait le plus fort taux de classement (7,9% de succès, 6,4% d'échecs) contrairement au samedi concentrant le maximum d'appels (5,6% de succès et 6,2% d'échecs par tentative d'appel).

Les appels passés en fin de journée (entre 18h et 20h45) permettaient de classer le maximum de numéros. Ceux passés entre 19h et 20h aboutissaient à 8,9% de succès et 7,9% d'échec. On observe ici encore un déphasage entre le pic des appels émis (entre 17h et 18h) et le pic de classement (entre 18h et 20h).

Toutes les tendances précédentes sont observées à la fois sur les téléphone fixes et sur les portables. Elles conduisent à se poser la question de l'organisation des moyens de télé-enquêteurs dans la journée et dans la semaine.

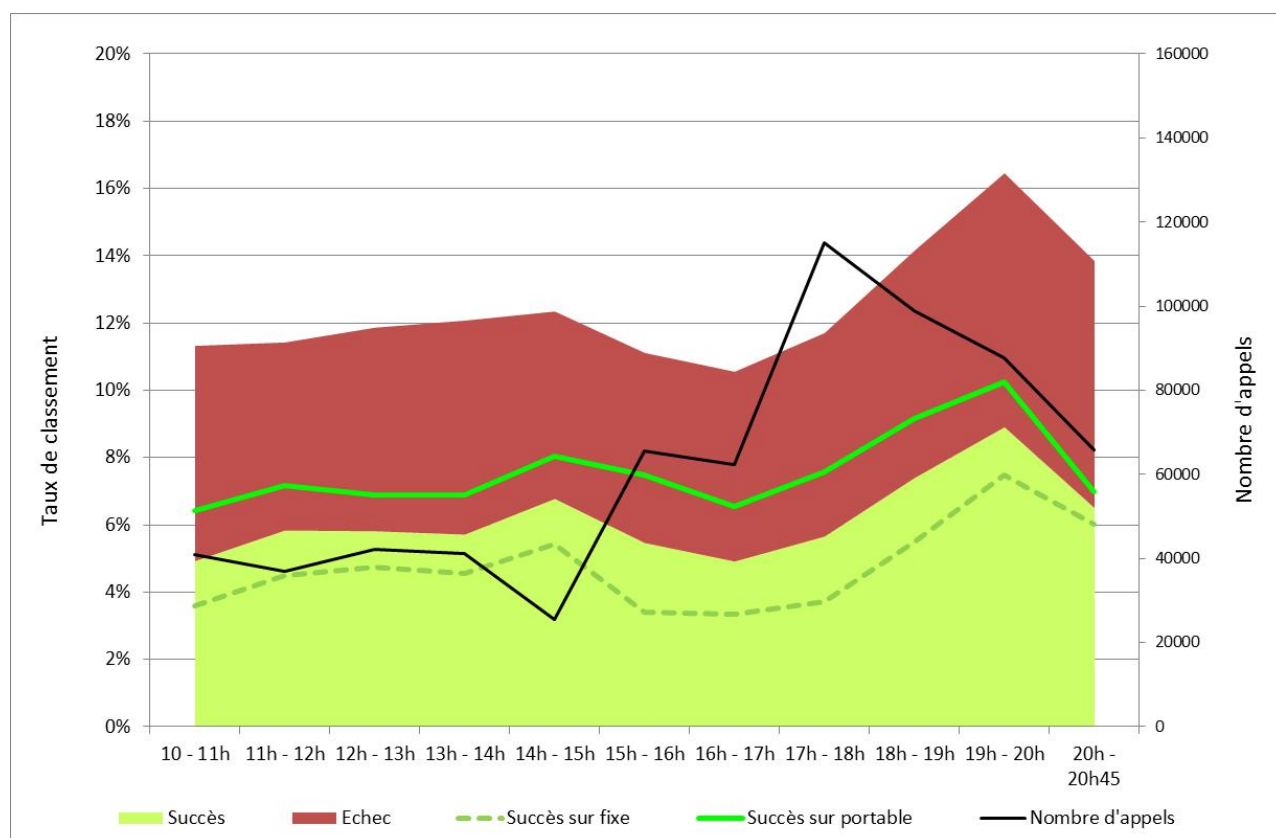
Graphique 5. Classement des appels en fonction du jour de l'appel



Source : enquête 2010 auprès de la Génération 2007

Champ : ensemble des tentatives d'appels effectuées auprès des 137 500 individus directement envoyés en enquête

Graphique 6. Classement des appels en fonction de l'heure de l'appel



Source : enquête 2010 auprès de la Génération 2007

Champ : ensemble des tentatives d'appels effectuées auprès des 137 500 individus directement envoyés en enquête

2.2.1. Classement final des numéros

On s'intéresse dans ce paragraphe uniquement aux numéros de téléphone pour lesquels au moins tentative d'appel a été effectuée. Le classement définitif d'un numéro est déterminé par le résultat de la dernière tentative d'appel. En moyenne, les numéros de téléphone classés en succès ont été appelés 3,5 fois et ceux classés en échec 2,9 fois.

Les numéros de téléphone portable issus de la base de sondage ainsi que les numéros issus des recherches téléphoniques de la phase A ont donné lieu à des succès dans plus de 30% des cas. Les numéros de téléphone fixe issus de la base de sondage ont permis d'aboutir à un succès dans 24% des cas. Pour les numéros de téléphone issus des recherches de coordonnées téléphoniques élargies, le taux de réussite est plus faible. Il atteint cependant 16 % pour les individus retrouvés sur les nom – prénom – département.

Tableau 6. Classement du dernier appel selon l'origine du numéro

Origine du numéro de téléphone	Succès	Echec	Non classé
Etablissement de formation : fixe	24%	24%	52%
Etablissement de formation : portable	31%	29%	40%
Recherche phase A, hors individu déménagé	34%	27%	39%
Recherche phase A sur individu déménagé	31%	24%	45%
Recherche manuelle Céreq	30%	31%	39%
Recherche élargie : Nom - prénom - département	16%	31%	53%
Recherche élargie : Nom - prénom -région	8%	44%	48%
Recherche élargie : Nom - commune	8%	59%	33%
Recherche élargie : Nom - prénom - Ile de France	3%	47%	50%
Recherche élargie : Nom - département	5%	61%	34%
Total	28%	28%	43%

Source : enquête 2010 auprès de la Génération 2007

Champ : ensemble des 151 800 numéros de téléphone exploités pour les individus directement envoyés en enquête

2.3. Téléphones non utilisés

A l'issue de la collecte, parmi les individus directement envoyés en enquête, 43 500 ont été classés en succès (31,6%), 24 400 ont été classés en échec (17,8%) et 69 600 (50,6%) n'ont pas pu être définitivement classés.

Les individus non définitivement classés à la fin de l'enquête disposaient de 152 400 numéros de téléphone, dont 76 500 seulement ont été utilisés. En fin de collecte, il subsistait 38 100 individus (dont 12 800 issus de l'échantillon de réserve) pour lesquels au moins un numéro de téléphone n'avait jamais été exploité. Pour 19 700 d'entre eux (dont 7 700 issus de l'échantillon de réserve), au moins un numéro de téléphone issu de la base de sondage ou des recherches téléphoniques de la phase A n'a pas été utilisé.

Tableau 7. Utilisation des numéros de téléphone des individus non classés à la fin de la collecte

Origine du numéro de téléphone	Nombre total de numéros disponibles	Nombre de numéros non exploités	Nombre d'individus présentant au moins un numéro de l'origine indiquée non exploité
Etablissement de formation : téléphone fixe	35 400	14 200	14 000
Etablissement de formation : téléphone portable	35 500	-	-
Recherche phase A (hors individu ayant déménagé) ou recherche manuelle Cereq	21 800	6 700	6 700
Recherche phase A sur individu ayant déménagé	1 400	600	600
Recherche élargie : Nom - prénom - département	3 700	3 400	2 400
Autre type de recherche élargie	54 600	51 000	23 300
Total	152 400	75 900	

Source : Enquête 2010 auprès de la Génération 2007

Champ : ensemble des numéros de téléphone disponibles pour les 69 600 individus directement envoyés en enquête et non classés à l'issue de la collecte

Lecture : Pour les individus directement envoyés en enquête et non classés à l'issue de la collecte, 35 400 numéros de téléphone fixe avaient été transmis par l'établissement de formation, dont 14 200 ont été exploités par le prestataire. 14 000 individus possédaient au moins un numéro de fixe transmis par l'établissement de formation et non exploité.

3. Rang d'appel, type et origine du numéro et caractéristiques des répondants

Dans cette dernière partie, nous restreignons l'analyse aux 23 500 individus directement envoyés en enquête, dans le champ retenu par le Céreq et qui ont répondu à l'intégralité du questionnaire. Parmi eux, 11 500 ont répondu sur un téléphone portable et 12 000 sur un téléphone fixe.

Dans la suite des travaux de Lopez et Sigot (2004) portant sur l'enquête 2001 auprès de la « Génération 1998 », différentes régressions logistiques ont été testées sur l'enquête 2010 pour modéliser la probabilité d'être en emploi, au chômage, de vivre en couple à la date de l'enquête. Lopez et Sigot avaient mis en évidence des biais liés à une insistance insuffisante. Ils ont ainsi montré, grâce à une modélisation logistique intégrant les variables explicatives classiques, que la probabilité d'être en emploi à la date de l'enquête augmentait avec le nombre d'appels nécessaires pour joindre l'individu.

Les variables d'intérêt sont celles liées à la gestion de la collecte (nombre de tentatives d'appels avant de réaliser le questionnaire, type et origine du numéro de téléphone). Des variables de contrôle sont également intégrées pour isoler leurs effets spécifiques. Il s'agit de variables classiques, telles que le type de formation suivie, le plus haut diplôme obtenu, la région de l'établissement de sortie, le genre ou le mode d'habitat.

3.1. Emploi à la date d'enquête

Soulignons tout d'abord que les effets associés aux régions et niveau de diplôme vont dans le sens attendu : on retrouve les différences de situation régionales du marché du travail (sachant que les individus formés dans une région ont de grandes chances de débiter leur vie active dans cette même région). L'effet favorable et croissant du niveau de diplôme sur les conditions d'emploi se retrouve nettement.

Concernant les variables d'intérêt, plus le nombre de tentatives d'appels effectuées sur un numéro de téléphone donné avant de réaliser l'enquête est élevé, plus l'individu contacté a de chances d'être en emploi (Odds Ratio associé à une tentative d'appel supplémentaire = 1,06). Le type de numéro de téléphone utilisé est également lié à la probabilité d'être en emploi, les personnes contactées sur un téléphone fixe se révélant moins souvent en emploi que celles contactées sur un portable (OR=0,86). L'origine du numéro de téléphone a également son importance puisque les individus contactés sur un numéro de téléphone issu des recherches phase A a plus de chances d'être en emploi qu'un individu contacté sur un numéro issu d'une recherche de coordonnées élargie (OR=1,28).

Le jour de la semaine où a été réalisée l'enquête a été testé dans le modèle mais n'a pas d'incidence significative sur la probabilité d'être en emploi.

Des régressions logistiques indépendantes réalisées sur les téléphones portables et fixes montrent qu'un rang d'appel élevé est associé à une probabilité d'être en emploi plus forte uniquement sur les téléphones fixes. Sur les portables, il n'existe pas de relation significative. Cette distinction est peut-être liée au fait que les personnes en emploi sont joignables sur un téléphone fixe sur des plages horaires plus restreintes que sur un téléphone portable. L'accroissement de la part des numéros de téléphones portables dans la base de sondage entre « Génération 2004 » et « Génération 2007 » a donc pu contribuer à réduire le biais lié à une faible insistance téléphonique sur l'estimation du taux d'emploi.

Tableau 8. Estimation de la probabilité d'être en emploi à la date de l'enquête

Variable	Prob >Khi 2	Odds Ratio	Intervalle de confiance de Wald à 95 %	
Nombre de tentatives d'appels	<,0001	1,06	1,04	1,07
Type de numéro	0,002			
Portable		Réf	-	-
Fixe		0,86	0,78	0,94
Origine du numéro	0,006			
Recherche élargie		Réf		
Etablissement de formation		1,10	0,84	1,44
Recherche phase A, recherche manuelle		1,28	0,98	1,68
Formation suivie en 2007	<,0001			
Bac + 1 à bac + 5 par apprentissage		Réf		
Collège		0,45	0,29	0,70
CAP/BEP apprentissage		0,82	0,61	1,11
CAP/BEP scolaire		0,57	0,42	0,77
Lycée		0,70	0,54	0,92
Bac + 1 à bac + 2		0,46	0,36	0,59
Formation de la santé et du social		2,83	2,08	3,86
Licence professionnelle		0,88	0,61	1,25
Licence université		0,28	0,21	0,38
Master 1		0,44	0,32	0,59
IUFM		1,29	0,89	1,87
Formation de l'animation et du sport		0,92	0,66	1,28
Formations bac+5 et plus		0,72	0,52	0,98
Plus haut diplôme obtenu	<,0001			
Non diplômé		Réf		
CAP-BEP-MC tertiaire		1,91	1,60	2,29
CAP-BEP-MC industriel		2,63	2,22	3,12
Bac général ou technologique		1,75	1,42	2,15
Bac pro tertiaire		2,47	2,01	3,05
Bac pro industriel		3,84	3,14	4,69
Bac+2 tertiaire		4,96	3,95	6,24
Bac+2 industriel		4,97	3,88	6,37
2ème cycle		3,43	2,63	4,47
3ème cycle ou école		4,09	3,00	5,59
Région de l'établissement de formation	0,004			
Régions Haute-Normandie et Bretagne		Réf		
Région Ile de France		1,13	0,97	1,31
Régions Basse-Normandie, Alsace et Lorraine		1,08	0,91	1,28
Région Nord Pas de Calais		0,94	0,82	1,08
Région Rhône-Alpes		1,10	0,93	1,29
Région Languedoc-Roussillon		0,86	0,69	1,07
Régions PACA et Corse		0,87	0,72	1,05
Régions Bourgogne, Centre, Pays de Loire, Auvergne et		1,17	1,00	1,36
Régions Champagne-Ardenne et Picardie		0,95	0,76	1,19
Régions franche-Comté, Aquitaine, Midi-Pyrénées, Poitou-		0,94	0,85	1,05
Lieu de résidence de l'individu en 2007	0,001			
Habite en ZUS		Réf		
N'habite en ZUS		1,23	1,08	1,41
Information non disponible		1,05	0,88	1,24
Genre	<,0001			
Homme		Réf		
Femme		0,725	0,672	0,784
Mode d'habitat en 2010	<,0001			
Seul		Réf		
En couple		0,52	0,48	0,57
Chez les parents		1,19	1,08	1,31

Source : enquête 2010 auprès de la Génération 2007

Champ : enquêtes réalisées dans le champ de l'enquête Céreq pour les individus directement envoyés en enquête, associées au numéro de téléphone ayant permis la réalisation de l'enquête

3.2. Chômage à la date de l'enquête

Les variables influençant la probabilité d'être au chômage interviennent de manière symétrique sur la probabilité d'être en emploi. Plus le nombre de tentatives d'appels effectués sur un numéro de téléphone donné avant de réaliser l'enquête est élevé, moins l'individu contacté a de chances d'être au chômage (Odds Ratio associé à une tentative d'appel supplémentaire = 0,92). Le type de numéro de téléphone utilisé est également lié à la probabilité d'être en emploi, les personnes contactées sur un téléphone fixe se révélant plus souvent au chômage que celles contactées sur un portable (OR=1,24). Enfin, les individus contactés sur un numéro de téléphone issu des recherches phase A ont moins de chances d'être au chômage que ceux contactés sur un numéro issu d'une recherche de coordonnées élargie (OR=0,74).

Des régressions logistiques indépendantes réalisées sur les téléphones portables et fixes montrent qu'un rang d'appel élevé est associé à une probabilité d'être au chômage à la fois sur les téléphones fixes et sur les portables.

Tableau 9. Estimation de la probabilité d'être au chômage à la date de l'enquête

Variable	Prob >Khi 2	Odds Ratio	Intervalle de confiance de Wald à 95 %	
Nombre de tentatives d'appels	<,0001	0,92	0,91	0,94
Type de numéro	0,000			
Portable		Réf	-	-
Fixe		1,24	1,10	1,39
Origine du numéro	0,008			
Recherche élargie		Réf		
Etablissement de formation		0,88	0,64	1,20
Recherche phase A, recherche manuelle		0,74	0,54	1,02
Formation suivie en 2007	<,0001			
Formation de la santé et du social		Réf		
Collège		4,32	2,61	7,16
CAP/BEP apprentissage		4,17	2,96	5,87
CAP/BEP scolaire		5,29	3,75	7,47
Lycée		4,11	3,01	5,62
Bac + 1 à bac + 2		5,40	4,06	7,19
Bac + 1 à bac + 5 par apprentissage		3,83	2,62	5,62
Licence professionnelle		4,75	3,05	7,41
Licence université		6,95	4,74	10,18
Master 1		7,80	5,27	11,56
IUFM		2,15	1,29	3,59
Formation de l'animation et du sport		3,07	2,05	4,60
Formations bac+5 et plus		4,83	3,23	7,23
Plus haut diplôme obtenu	<,0001			
Non diplômé		Réf		
CAP-BEP-MC tertiaire		0,60	0,49	0,73
CAP-BEP-MC industriel		0,46	0,38	0,55
Bac général ou technologique		0,35	0,28	0,45
Bac pro tertiaire		0,48	0,38	0,61
Bac pro industriel		0,30	0,24	0,38
Bac+2 tertiaire		0,25	0,19	0,33
Bac+2 industriel		0,24	0,18	0,32
2ème cycle		0,25	0,18	0,35
3ème cycle ou école		0,31	0,21	0,46
Région de l'établissement de formation	0,002			
Régions Haute-Normandie et Bretagne		Réf		
Région Ile de France		0,90	0,75	1,08
Régions Basse-Normandie, Alsace et Lorraine		0,99	0,81	1,22
Région Nord Pas de Calais		1,26	1,07	1,48
Région Rhône-Alpes		0,91	0,74	1,12
Région Languedoc-Roussillon		1,29	0,98	1,69
Régions PACA et Corse		1,05	0,83	1,32
Régions Bourgogne, Centre, Pays de Loire, Auvergne et		0,94	0,78	1,13
Régions Champagne-Ardenne et Picardie		1,26	0,96	1,64
Régions franche-Comté, Aquitaine, Midi-Pyrénées, Poitou-		1,08	0,95	1,24
Lieu de résidence de l'individu en 2007	<,0001			
Habite en ZUS		Réf		
N'habite en ZUS		0,74	0,64	0,87
Information non disponible		0,90	0,74	1,10
Genre	0,01			
Homme		Réf		
Femme		1,127	1,026	1,238
Mode d'habitat en 2010	<,0001			
Seul		Réf		
En couple		1,06	0,94	1,21
Chez les parents		2,40	2,12	2,71

Source : enquête 2010 auprès de la Génération 2007

Champ : enquêtes réalisées dans le champ de l'enquête Céreq pour les individus directement envoyés en enquête, associées au numéro de téléphone ayant permis la réalisation de l'enquête

3.3. Autres influences du rang d'appel

Des modèles logistiques faisant intervenir les variables présentées précédemment visaient à déterminer s'il existait un lien entre le rang de l'appel et les caractéristiques de l'insertion. Le rang d'appel n'est pas lié de manière significative à la probabilité d'être actif à la date de l'enquête ou au sexe de la personne interrogée. Pour les personnes en emploi, il n'est pas lié à la probabilité d'être en emploi à durée déterminée.

Par contre, plus le rang d'appel est faible, plus la probabilité que l'enquêté vive chez ses parents à la date de l'enquête est importante. Enfin, le rang d'appel est plus élevé chez les sortants de l'enseignement supérieur que chez les sortants du secondaire.

Conclusion

Afin d'optimiser les chances de joindre les individus échantillonnés pour l'enquête réalisée auprès de la « Génération 2007 », le Céreq a demandé aux établissements de formation, lors de la constitution de la base de sondage, de transmettre les numéros de téléphone dont ils disposaient. Il a ensuite réalisé des recherches automatiques de coordonnées téléphoniques dans les fichiers de France Télécom. Enfin, pour un petit nombre d'individus, il a effectué des recherches manuelles sur les pages blanches. Afin d'optimiser les chances de joindre un individu sur un numéro de téléphone donné, il a demandé au prestataire de réaliser un minimum de dix tentatives d'appels infructueuses avant d'abandonner le numéro.

En pratique, le brassage du fichier n'a pas été satisfaisant. D'une part, le premier appel sur le numéro prioritaire a été souvent tardif : après six semaines d'enquêtes, aucune tentative d'appel n'avait été effectuée sur le numéro de priorité 1 pour la moitié des individus directement envoyés en enquête. D'autre part, la fréquence des rappels a été insuffisante. Il s'en est suivi en fin de collecte un abandon très élevé des appels sur un numéro donné avant les dix tentatives préconisées. Cela est fortement regrettable, au regard tout d'abord du taux de réponse global, impacté de manière négative par ces abandons précoces. Cet état de fait a induit des biais, eu égard notamment aux téléphones fixes pour lesquels nous avons montré que la probabilité d'être en emploi augmentait avec le rang de l'appel permettant de réaliser le questionnaire tandis que celle d'être au chômage diminuait. Les biais semblent plus limités pour les portables.

Le choix d'une utilisation séquentielle des numéros de téléphone a par ailleurs conduit à ne pas exploiter du tout un nombre important de numéros pour des individus pour lesquels des tentatives d'appels restaient envisageables.

L'analyse des résultats d'appels selon l'origine du numéro de téléphone confirme la hiérarchie des numéros effectuée par le Céreq. Les numéros de téléphone portable issus des établissements de formation donnent les meilleurs résultats tandis que les numéros issus des recherches de coordonnées téléphoniques élargies, mis à part celles sur le nom-prénom-département, présentent un très faible rendement.

Quelques suggestions pour les prochaines enquêtes peuvent être avancées. Les recherches téléphoniques phase A, qui ont montré leur pertinence, ainsi que les recherches élargies sur les nom-prénom-département, devraient être poursuivies. Pour les individus ne disposant pas de coordonnées téléphoniques à l'issue des recherches phase A, il serait intéressant de réaliser des recherches manuelles sur les pages blanches. Cela nécessiterait des moyens humains importants mais permettrait de ne pas se priver d'une partie des individus pour lesquels les recherches sont infructueuses. Les recherches élargies (hors nom-prénom-département) pourraient être réalisées uniquement sur les individus ne disposant d'aucune coordonnée téléphonique à l'issue des étapes précédentes.

Le brassage du fichier par le prestataire devrait être plus important, notamment en début d'enquête. Le délai entre les tentatives d'appels pourrait par exemple être raccourci. Au vu des résultats, un ajustement du dimensionnement du plateau selon les jours et les plages horaires peut également être envisagé. L'exploitation simultanée de plusieurs numéros pour un même individu, par exemple à partir du milieu de collecte, permettrait d'utiliser l'ensemble des numéros. Enfin, la réalisation effective d'au moins dix tentatives d'appels est primordiale, tant pour augmenter le taux de réponse que pour limiter les biais. Le nombre de dix tentatives est un minimum, celui de 15 paraissant souhaitable.

Bibliographie

- [1] Lopez A., Sigot J.C., « Analyse de deux risques de biais dans une enquête téléphonique partant d'une base de sondage nominative » dir. Pascal Ardilly, DUNOD 2004.
- [2] d'Haultefoeuille X., Fevrier P., « Persévérance, acceptabilité et acceptation : une nouvelle modélisation de la non-réponse », Journées de Méthodologie Statistique, décembre 2002.
- [3] Duplouy B., « Améliorer le taux de réponse dans les enquêtes téléphoniques : une expérimentation de recherches de coordonnées téléphoniques sur l'enquête Génération 2004 », Journées de Méthodologie Statistique, mars 2009.