

## Enquête « Génération 2004 » Méthodologie et bilan

1ère interrogation - printemps 2007

**Christel Aliaga**

**Bérangère Duploux**

**Stéphane Jugnot**

**Pascale Rouaud**

**Florence Ryk**

Céreq

10, place de la Joliette BP21321

13 567 Marseille Cedex 02

Pascale Rouaud et Stéphane Jugnot ont coordonné la rédaction de ce document.

Il a également bénéficié des travaux de Nathalie Marchal, Mickaele Molinari et Jean-Claude Sigot.

Ce document est présenté sur le site du Céreq afin de favoriser la diffusion et la discussion de résultats de travaux d'études et de recherches. Il propose un état d'avancement provisoire d'une réflexion pouvant déboucher sur une publication. Les hypothèses et points de vue qu'il expose, de même que sa présentation et son titre, n'engagent pas le Céreq et sont de la responsabilité des auteurs.

**mai 2010**



## **SYNTHÈSE**

---

Cette note propose une présentation méthodologique de l'enquête « Génération 2004 », collectée par téléphone en 2007 après des jeunes sortis de formation initiale trois ans auparavant.

Les enquêtes « Génération », mises en place par le Céreq, forment un dispositif d'enquêtes régulières auprès des jeunes sortants de formation initiale. Elles étudient l'insertion des jeunes sur le marché du travail en fonction de leur niveau de formation, de leur diplôme et de leur spécialité. Tous les trois ans, une nouvelle cohorte de jeunes sortis de formation est ainsi interrogée, trois ans après leur sortie, sur leurs premiers pas sur le marché du travail.

Destinée principalement aux utilisateurs des données de l'enquête, ainsi qu'aux personnes intéressées plus généralement par les dimensions méthodologiques, cette note aborde notamment la structure du questionnaire, le protocole de collecte, la constitution de la base de sondage, les taux de succès et le traitement de la non réponse.



## SOMMAIRE

---

<b>1. PRÉSENTATION DE L'ENQUÊTE « GÉNÉRATION 2004 »</b> .....	<b>7</b>
<b>1.1. Le dispositif des enquêtes « Génération »</b> .....	<b>7</b>
<b>1.2. Le champ de l'enquête</b> .....	<b>8</b>
<b>1.3. Les extensions</b> .....	<b>9</b>
<b>1.4. Les questionnaires</b> .....	<b>10</b>
1.4.1. Le questionnaire complet .....	12
1.4.2. Le questionnaire allégé .....	14
1.4.3. Zoom sur le calendrier professionnel.....	14
1.4.4. Les extensions de questionnement .....	16
<b>1.5. Le calendrier de l'enquête</b> .....	<b>17</b>
<b>2. LA CONSTITUTION DE LA BASE DE SONDAGE</b> .....	<b>19</b>
<b>2.1. La collecte auprès des établissements</b> .....	<b>19</b>
2.1.1. Collecte des bases rectorales.....	19
2.1.2. Collecte auprès des autres établissements.....	20
<b>2.2. L'apurement de la base de sondage</b> .....	<b>21</b>
<b>3. LA CONSTITUTION DE L'ÉCHANTILLON</b> .....	<b>24</b>
<b>3.1. Le cadre général du plan de sondage</b> .....	<b>24</b>
<b>3.2. L'estimation du taux de couverture des individus entrant dans le champ (P1)</b> .....	<b>26</b>
<b>3.3. L'estimation du taux de non réponse prévisible des individus (P2)</b> .....	<b>29</b>
<b>3.4. L'estimation du coefficient d'échantillonnage (P3)</b> .....	<b>30</b>
3.4.1. Première étape : le tirage à blanc d'un échantillon hors extension .....	30
3.4.2. Deuxième étape : le calcul des coefficients de suppléments de tirage.....	30
<b>3.5. L'affectation des personnes échantillonnées au questionnaire « complet »     ou au questionnaire « allégé »</b> .....	<b>31</b>
<b>3.6. L'affectation des personnes échantillonnées à l'échantillon principal ou à la réserve     d'échantillon</b> .....	<b>32</b>
<b>4. LA COLLECTE</b> .....	<b>34</b>
<b>4.1. L'enrichissement des coordonnées téléphoniques</b> .....	<b>34</b>
4.1.1. Le protocole de recherche .....	35
4.1.2. Bilan de l'enrichissement.....	35
<b>4.2. La phase de pré qualification</b> .....	<b>38</b>
4.2.1. La détermination des individus concernés par la phase de pré qualification .....	38
4.2.2. Le déroulement de la phase de pré qualification.....	40
4.2.3. Bilan de la phase de pré qualification .....	41

<b>4.3. L'amélioration des phases d'enrichissement et de pré qualification par rapport aux enquêtes précédentes .....</b>	<b>43</b>
4.3.1. Nature des évolutions.....	43
4.3.2. Le bilan de deux tests préalables.....	45
<b>4.4. La collecte .....</b>	<b>45</b>
4.4.1. L'organisation générale de la collecte.....	45
4.4.2. Le suivi de la collecte.....	48
4.4.3. Le taux de réponse final .....	48
4.4.4. Autres enseignements de la collecte .....	50
<b>5. LES TRAITEMENTS EN AVAL (APUREMENT, CODAGE, REpondÉRATION) .....</b>	<b>52</b>
<b>5.1. Les tables finales.....</b>	<b>52</b>
5.1.1. Les trois tables associées aux questionnaires complets .....	52
5.1.2. Les trois tables associées aux questionnaires allégés.....	53
5.1.3. Les trois tables associées à chaque extension de champ.....	54
<b>5.2. Des fichiers bruts aux fichiers finaux .....</b>	<b>54</b>
5.2.1. La construction des bases brutes .....	54
5.2.2. La vérification des filtres .....	56
5.2.3. La suppression d'observations non exploitables.....	56
5.2.4. Le traitement des salaires.....	57
5.2.5. L'enrichissement par des informations de la base de sondage.....	57
5.2.6. Le repérage du plus haut diplôme obtenu .....	57
5.2.7. Le codage des spécialités de formation.....	58
5.2.8. Le codage du secteur d'activité des périodes d'emploi.....	58
5.2.9. Le codage de la profession des périodes d'emploi.....	59
5.2.10. Le codage du contrat de travail.....	60
5.2.11. Les variables de localisation géographique .....	61
5.2.12. Le codage des réponses en clair.....	62
5.2.13. Le contrôle de cohérence finale .....	62
5.2.14. L'anonymisation des fichiers.....	63
<b>5.3. La pondération finale.....</b>	<b>63</b>
5.3.1. La prise en compte de la non réponse globale .....	65
5.3.2. Le calage sur marges.....	67
<b>6. BILAN ET PERSPECTIVES POUR LES ENQUÊTES FUTURES .....</b>	<b>68</b>
<b>ANNEXE 1 : COMPLÉMENTS SUR LES POIDS DE TIRAGE .....</b>	<b>71</b>
<b>ANNEXE 2 : BILAN DES TESTS DE QUESTIONNAIRE.....</b>	<b>78</b>
<b>ANNEXE 3 : OPTIMISER LE NOMBRE D'APPEL .....</b>	<b>81</b>
<b>ANNEXE 4 : LES OUTILS DE LA SENSIBILISATION DES JEUNES .....</b>	<b>91</b>
<b>ANNEXE 5 : REGROUPEMENT DE LA NES EN 114 POSITIONS EN 67 POSTES.....</b>	<b>99</b>
<b>ANNEXE 6 : LE REDRESSEMENT DE LA NON REponse TOTALE .....</b>	<b>101</b>

<b>ANNEXE 7 : UNE EXPÉRIMENTATION DE RECHERCHE DE COORDONNÉES TÉLÉPHONIQUES POUR AMÉLIORER LE TAUX DE RÉPONSE A L'ENQUÊTE ..</b>	<b>115</b>
<i>1.1. Historique du taux d'équipement .....</i>	<b>117</b>
<i>1.2. Les atouts des enquêtes téléphoniques.....</i>	<b>118</b>
<i>1.3. Difficultés et limites des enquêtes téléphoniques .....</i>	<b>119</b>
<i>2.1. Présentation de l'enquête Génération 2004 du Céreq .....</i>	<b>122</b>
<i>2.2. La base de sondage de Génération 2004.....</i>	<b>124</b>
2.2.1. Des coordonnées non exhaustives.....	124
2.2.2. La mobilité géographique des individus .....	125
<i>3.1. L'enrichissement des coordonnées téléphoniques dans l'enquête « Génération 98 ».....</i>	<b>127</b>
<i>3.2. Une expérience niçoise .....</i>	<b>129</b>
<i>3.3. Le protocole de Génération 2004 .....</i>	<b>130</b>
<i>4.1. Un gain net de répondants .....</i>	<b>133</b>
<i>4.2. Des profils d'individus qui diffèrent .....</i>	<b>134</b>
<i>4.3. Quel redressement ? .....</i>	<b>139</b>
<i>4.4. Un coût important .....</i>	<b>140</b>
<b>INDEX DES SIGLES .....</b>	<b>147</b>



# 1. PRÉSENTATION DE L'ENQUÊTE « GÉNÉRATION 2004 »

---

## 1.1. Le dispositif des enquêtes « Génération »

Les enquêtes « Génération » s'intéressent à l'insertion et au cheminement des sortants du système éducatif lors de leurs premières années de vie active. Elles ont pour objectifs principaux de produire des indicateurs d'insertion (taux d'emploi, taux de chômage, taux d'emploi à durée indéterminée, etc.), selon les niveaux de formation, les filières, les spécialités, à destination des acteurs publics et sociaux. Elles proposent ainsi des informations qui contribuent à la compréhension des processus d'insertion et des différenciations des parcours en début de carrière.

### Encadré 1 - Le centre d'études et de recherches sur les qualifications

Créé au début des années 1970, devenu établissement public en 1985, le Centre d'études et de recherches sur les qualifications (Céreq) est placé sous la tutelle des ministères en charge de l'Education nationale et du Travail.

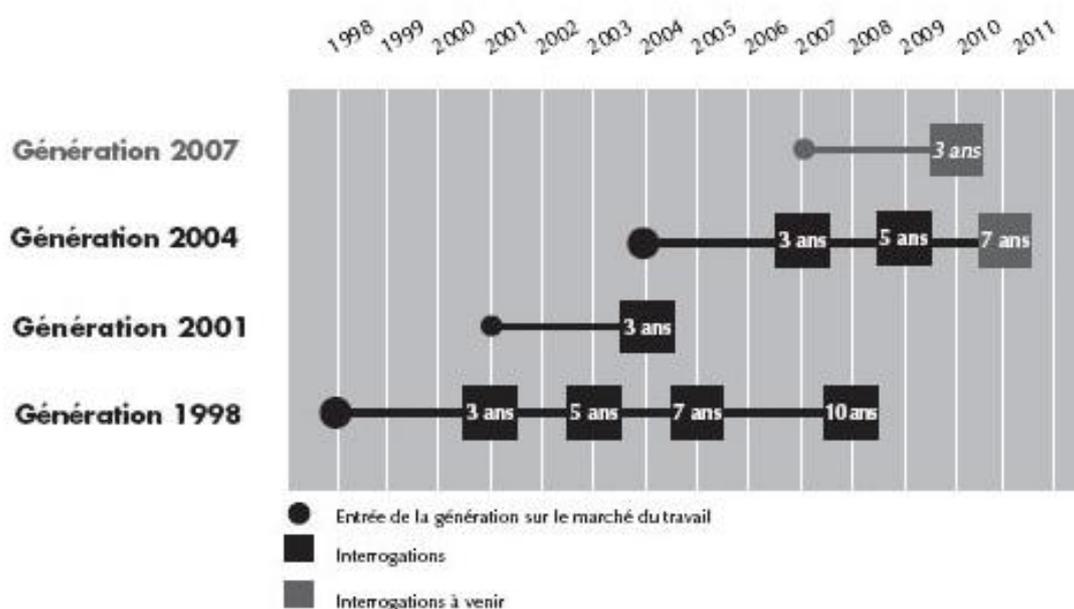
Il a pour missions de développer des études et des recherches, de collecter et d'exploiter des données originales dans le domaine de la relation formation-emploi, et de formuler des avis et propositions destinés à éclairer les choix en matière de politiques de formation.

La première enquête « Génération » a été effectuée en 1997 auprès de jeunes sortis du système éducatif en 1992 et issus de tous les niveaux de formation. Un dispositif régulier d'interrogation a été mis en place à partir de l'enquête effectuée en 2001 auprès de sortants du système éducatif en 1998 : les enquêtes s'enchaînent au rythme d'une génération nouvelle de sortants interrogée tous les trois ans et avec une alternance entre une enquête « génération pleine » et une « mini-génération ».

L'enquête « génération pleine » comprend plusieurs interrogations. La première interrogation, trois ans après la sortie du système éducatif, est principalement tournée vers la production d'indicateurs et des usages institutionnels. Les interrogations suivantes, à cinq et sept ans, sont plutôt centrées sur les usages analytiques (notamment sur la question des parcours, des mobilités sur moyen terme). Ces ré-interrogations permettent aussi d'approfondir certains constats issus de l'exploitation de la première interrogation (par exemple sur les insertions les plus problématiques).

L'enquête « mini-génération » se limite à la première interrogation à trois ans, avec une taille d'échantillon plus réduite. Cette enquête a vocation à réactualiser les indicateurs d'insertion sur les trois premières années d'insertion selon une grille d'analyse moins fine que pour les « générations pleines ».

**Figure 1. Enchaînement des enquêtes « Génération »**



Le cœur de l'enquête, pour la première interrogation comme pour les autres quand il y a lieu, est constitué par le calendrier professionnel qui permet de suivre mois par mois la situation des jeunes à l'issue de leur formation initiale, et de décrire les séquences successives d'emploi et de non emploi.

Les enquêtes « Génération » permettent aussi de répondre à des demandes d'extensions nationales ou régionales portant sur les jeunes issus de certains niveaux ou spécialités de formations ou sur les jeunes ayant bénéficié de certaines mesures pour la formation. Les échantillons peuvent donc être élargis en conséquence.

**Figure 2. Nombre de répondants à la première interrogation**

	Année de la première interrogation (trois après la sortie)	Nombre de répondants	Dont liés aux extensions d'échantillon
« Génération 1992 »	1997	26 000	
« Génération 1998 »	2001	55 000	19 000
« Génération 2001 »	2004	25 000	15 000
« Génération 2004 »	2007	65 000	37 000

## 1.2. Le champ de l'enquête

L'enquête « Génération 2004 » concerne les « primo sortants » de formation initiale en 2003-2004 (année scolaire). Les sortants de formation qui avaient déjà interrompu leurs études au moins un an avant l'année scolaire considérée sont hors champ. Tous les niveaux et domaines de formations sont concernés.

De façon plus précise, les critères d'éligibilité pour être dans le champ retenu, nommé ensuite « champ Céreq », sont les suivants :

- avoir été inscrit dans un établissement de formation en France métropolitaine durant l'année scolaire 2003-2004 ;

- avoir quitté le système éducatif entre octobre 2003 et octobre 2004<sup>1</sup> ;
- ne pas avoir interrompu ses études durant une année ou plus avant l'année scolaire 2003-2004 (sauf pour raison de santé) ;
- ne pas avoir repris ses études pendant l'année qui a suivi l'entrée sur le marché du travail ;
- avoir 35 ans ou moins en 2004 ;
- être localisé en France métropolitaine au moment de l'enquête (ce qui exclut donc notamment les personnes poursuivant des études à l'étranger ou y allant pour un premier emploi).

Ces conditions sont cumulatives.

Quelques points particuliers concernent l'application de ces critères :

- Bien qu'il s'agisse de contrats de travail au même titre que les contrats de professionnalisation ou que les contrats de qualification, les contrats d'apprentissage sont considérés comme relevant de la formation initiale. Une personne repérée comme sortant de formation en 2003-2004 qui poursuit par un contrat d'apprentissage en 2004-2005 est donc considérée en poursuite d'études, donc hors champ. Ce n'est pas le cas s'il s'agit d'un contrat de professionnalisation ou de qualification.
- Une personne sortie d'un établissement de formation en 2003-2004 qui poursuit des cours par correspondance ou des cours du soir en 2004-2005 est considérée comme en poursuite d'études, donc hors champ, si elle n'a pas d'emploi en parallèle. Si elle a un emploi en parallèle, sa situation d'emploi prime.
- Une personne sortie d'un établissement de formation en 2003-2004 mais poursuivant ses études en 2004-2005 comme élève fonctionnaire n'est pas considérée comme étant en poursuite d'études ; elle est considérée en emploi. Les personnes en deuxième année d'IUFM sont donc, en particulier, considérées comme étant en emploi.

Les différences de champ qui existent entre les trois cohortes (génération des sortants de 1998, 2001 et 2004) sont mineures : les sortants de classes de 6<sup>ème</sup> et 5<sup>ème</sup>, inclus dans le champ des enquêtes « Génération 1998 » et « Génération 2001 » sont exclus du champ de l'enquête « Génération 2004 » ; les étudiants étrangers de l'Université (repérables dans SISE, Système d'Information sur le Suivi de l'Étudiant) étaient hors champ pour les enquêtes « Génération 1998 » et « Génération 2001 » mais sont dans le champ de l'enquête « Génération 2004 ». Enfin, la période de référence pour la date de sortie était l'année civile pour les enquêtes « Génération 1998 » et « Génération 2001 » (janvier à décembre); c'est l'année scolaire pour l'enquête « Génération 2004 » (octobre 2003 à octobre 2004).

### **1.3. Les extensions**

Les acteurs intervenant dans le domaine de la formation sont multiples, avec leurs questionnements propres mais aussi avec un questionnaire commun sur le devenir de leur public cible. A la demande de certains d'entre eux, ministères, conseils régionaux ou observatoires de branches, des extensions sont régulièrement adossées aux enquêtes « Génération ». Elles permettent de réaliser des analyses ciblées sur des publics, des filières ou des territoires particuliers dans un contexte plus large, par rapprochement avec des indicateurs de référence nationaux.

---

<sup>1</sup> Sauf pour les extensions Sport et Animation, Santé Social et Thèses où on enquête les jeunes sortis entre octobre 2003 et décembre 2004.

Plusieurs types d'extensions sont possibles, éventuellement combinables :

- une extension d'échantillon, de façon à disposer d'un nombre de répondants suffisant pour permettre des analyses statistiquement pertinentes sur les catégories de sortants ciblées ;
- une extension de champ, pour interroger également sur les catégories de sortants ciblées les personnes non retenues dans le champ Céreq (par exemple en levant la restriction aux seuls primo sortants) ;
- une extension de questionnement, pour poser quelques questions complémentaires sur des catégories de sortants ciblées.

Pour la première interrogation de l'enquête « Génération 2004 », le nombre de répondants visés pour l'enquête nationale hors extensions était de 30 000, auquel s'ajoutaient 40 000 répondants attendus supplémentaires répartis entre seize extensions (voir figure 3).

#### **1.4. Les questionnaires**

Au regard du volume des extensions d'échantillon associées à l'enquête « Génération 2004 », passer l'intégralité du questionnaire du Céreq sur l'ensemble des personnes enquêtées, extensions comprises, ainsi que les extensions de questionnement auprès des personnes concernées, aurait pu rendre critique la capacité des prestataires à réaliser le nombre de questionnaires souhaité dans les délais programmés : il importe en effet que la période de collecte ne s'étale pas trop pour que la situation des personnes à la date d'enquête se réfère approximativement à la même date.

Par ailleurs, la plupart des demandeurs d'extensions souhaitaient disposer d'informations de cadrage cohérentes avec celles publiées par le Céreq dans ses analyses générales nationales, sans pour autant forcément aller dans le même degré de détail.

De ce fait, un questionnaire « allégé » a été mis en place pour certaines extensions, en supprimant certaines parties du questionnaire « complet » passé auprès de l'échantillon national du Céreq. A titre indicatif, la durée moyenne d'un questionnaire complet a été de vingt-six minutes, contre quatorze minutes pour le questionnaire allégé.

**Figure 3. Bilan des demandes d'extensions sur Génération 2004**

Institution ou organisme demandeur d'extension	Cible de sortants	Objectif de questionnaires supplémentaires	Type de questionnaire	Extension de champ	Extension de questionnement
Direction de l'Evaluation de la Prospective et de la Performance (DEPP)	Diplômés L3, M2 et Licence Professionnelle	11 000	Allégé	Non	Non
Directions Générales de l'Enseignement Supérieur et de la Recherche et de l'Innovation	Doctorats hors santé	1 300	Complet	Sorties jusqu'en décembre 2004	Module Thèse
Direction de la Recherche, des Etudes, de l'Evaluation et de la Statistique du Ministère de la Santé et des Affaires Sociales (Drees)	Formations sanitaires et sociales	4 800	Complet	Inclusion des post-initiaux, sorties jusqu'en décembre 2004, pas de limite d'âge	Module Post-initiaux
Ministère de la Culture	Formations de la Culture	850	Allégé	Non	Non
Institut Français de l'Environnement (Ifen)	Formations de l'environnement	1 300	Allégé	Non	Non
Ministère de la Jeunesse, des Sports et de la Vie Associative	Diplômés des formations du Sport	2 200	Allégé	Inclusion des post-initiaux, sorties jusqu'en décembre 2004, pas de limite d'âge	Modules Post-initiaux et Multi-activité
Association nationale pour la formation automobile (Anfa)	Formations de la filière automobile	2 100	Complet	Non	Non
Observatoire prospectif du commerce (FORCO)	Formations du commerce	1 550	Allégé	Non	Non
Groupe Statistique d'évaluation (GSE)	Tous les sortants (pour obtenir une représentativité au niveau régional sur un découpage homme/femme et secondaire/supérieur)	6 000	Allégé	Non	Non
Région Nord-Pas-de-Calais	Tous les sortants de Nord-Pas-de-Calais (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	750	Allégé	Non	Non
Région Provence-Alpes-Côte-d'Azur	Tous les sortants de Provence-Alpes-Côte-d'Azur (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	1 050	Allégé	Non	Non
Région Poitou-Charentes	Tous les sortants de Poitou-Charentes (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	1 800	Allégé	Non	Non
Région Lorraine	Tous les sortants de Lorraine (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	1 645	Allégé	Non	Non
Région Basse-Normandie	Tous les sortants de Basse-Normandie (pour obtenir une représentativité sur les niveaux 4, 5, 6)	1 800	Allégé	Non	Non
Direction de l'animation de la recherche, des études et des statistiques du ministère de l'emploi et des solidarités (Dares)	CFA		Complet	Non	Module Apprentissage
Association pour la gestion des fonds pour l'insertion professionnelle des personnes handicapées (Agefiph)	Tous les sortants		Complet	Non	Questions sur la santé et le handicap

\*10 niveaux de diplômes : Sans diplôme, CAP/BEP industriel, CAP/BEP tertiaire, Bac Général, Bac Industriel, Bac Tertiaire, Bac+2 Industriel, Bac+2 santé ou tertiaire, Second Cycle et Troisième Cycle ou Ecoles.

### ***1.4.1. Le questionnaire complet***

Le questionnaire complet commence par une partie filtre destinée à valider l'identification de l'individu et à vérifier les critères d'éligibilité : l'interlocuteur est-il le « bon individu » ? Fait-il partie du champ (champ Céreq ou champ d'une extension, selon les cas) ? Outre le nom et le prénom de la personne vérifiés à l'initialisation du contact, sont vérifiés le mois et l'année de naissance, ainsi que le nom de l'établissement de formation fréquenté en 2003-2004. Si la date de naissance et le nom de l'établissement diffèrent des informations contenues dans la base de sondage, l'interlocuteur n'est pas enquêté et on continue à rechercher l'individu cible. Si la différence ne porte que sur l'une ou l'autre des informations, l'individu est enquêté s'il appartient au champ de l'enquête.

Le questionnaire aborde ensuite successivement différents thèmes :

- *Le parcours scolaire* : motif de l'arrêt des études, retard ou non à l'entrée en 6<sup>e</sup>, nature de la classe de 3<sup>e</sup> et de la classe suivie après la troisième, orientation voulue ou subie, diplômes obtenus et spécialités, parcours résidentiel.
- *Les séjours à l'étranger dans le cadre de la scolarité*.
- *L'expérience du travail en cours de scolarité*, sous forme de stage ou d'emploi.
- *Le calendrier professionnel* : description des séquences d'emploi et de non emploi sur la base d'un calendrier mensuel, portant sur les trois années allant de la sortie de formation initiale jusqu'à la date d'enquête (voir 1.4.3).
- *Les perspectives professionnelles* : degré de satisfaction concernant la situation professionnelle actuelle ; souhait de progression et conditions (mobilité géographique, professionnelle, formation, etc.).
- *L'évolution du mode de cohabitation* : calendrier mensuel de mode de cohabitation (vivre chez ses parents/vivre en couple/vivre seul).
- *Les caractéristiques individuelles et la situation familiale* : pays de naissance de l'enquêté et de ses parents, nationalité des parents lors de l'enquête, situation d'activité et situation professionnelle des parents à la date de fin d'études, nombre d'enfants à la date d'enquête.
- *Le sentiment de discrimination* : existence d'une discrimination professionnelle ressentie depuis la fin des études et motif.

La logique de questionnement et les principales thématiques abordées sont communes aux questionnaires des enquêtes « Génération 1998 », « 2001 » et « 2004 » : description du parcours scolaire, calendrier professionnel et description des séquences, caractéristiques individuelles et familiales.

Néanmoins, des aménagements ont été introduits au fil des enquêtes et des spécificités existent pour chaque enquête. Les modules « thèse » et « séjours à l'étranger » ainsi que quelques items relatifs à la santé ont ainsi été introduits à partir de l'enquête « Génération 2001 ». Les questions sur la catégorie socioprofessionnelle ont été retravaillées et celles sur la situation du conjoint ont été supprimées. Pour l'enquête « Génération 2001 », il n'y a pas eu de calendrier habitat.

Pour l'enquête « Génération 2004 », le parcours scolaire et les stages en cours d'études sont plus détaillés que pour les enquêtes précédentes ; les modules « apprentissage » et « activité professionnelle » dans la séquence en entreprise actuelle sont apparus, les modules « séquences de non emploi » et « reprise d'études du passé » ont été raccourcis, les questions sur les perspectives professionnelles ont été retravaillées.

## **Encadré 2 - Une enquête de la statistique publique**

L'enquête « Génération 2004 », extensions comprises, est une enquête de la statistique publique et relève à ce titre de la loi n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Celle-ci définit le cadre de procédures destinées à garantir l'intérêt, la pertinence et la qualité des enquêtes publiques, ainsi que la confidentialité des informations collectées.

L'opportunité de l'enquête « Génération 2004 » a ainsi été discutée au Comité national de l'information statistique (Cnis), qui a émis un avis favorable en 2005 puis a attribué un label d'intérêt général et de qualité statistique à l'enquête (label n°2007X712AU du 20 novembre 2006).

### ***Une phase de concertation***

Au préalable, une phase de concertation a permis l'expression des attentes de diverses institutions, des partenaires sociaux et d'experts, à partir d'une présentation des bilans techniques et des utilisations des enquêtes précédentes.

Une première réunion, le 15 mars 2006, a rassemblé des représentants des différentes directions du ministère de l'Éducation nationale, du Conseil national de la formation professionnelle tout au long de la vie (CNFPTV), du Médef, de la Confédération française démocratique du travail (CFDT), de l'Union nationale des syndicats autonomes (UNSA), de l'Office national d'information sur l'enseignement et les professions (Onisep) et d'autres personnes qualifiées, notamment des chercheurs.

Une deuxième réunion, plus technique, le 17 mai 2006, a rassemblé des représentants de l'Administration (Direction générale de l'enseignement scolaire), de services statistiques ministériels (la Depp, Direction de l'évaluation, de la prospective et de la performance ; la Dares, Direction de l'animation de la recherche, des études et des statistiques), l'Insee, le laboratoire d'analyse sociologique et des méthodes appliquées aux sciences sociales (Lasmus), le laboratoire interdisciplinaire de recherche sur les ressources humaines et l'emploi (Lirhe), le Conseil de l'emploi, des revenus et de la cohésion sociale (Cerc) et l'observatoire régional de l'emploi, de la formation et des qualifications de Lorraine.

L'enquête a également été discutée au comité des tutelles du Céreq du 14 juin 2006.

### ***La confidentialité des réponses***

Les réponses apportées à l'enquête sont confidentielles. Les fichiers des réponses détaillées à l'enquête qui sont mis à disposition des chercheurs sont anonymes et ne comportent pas d'informations susceptibles de permettre une identification directe ou indirecte des personnes enquêtées. Par ailleurs, l'ensemble des traitements a fait l'objet de deux déclarations auprès de la Commission nationale informatique et libertés (Cnil) : une première portant sur la constitution de la base de sondage (déclaration n°1063222 sur la « constitution de la base nominative d'individus inscrits en formation initiale en 2004 » ; récépissé du 17 mars 2005) ; une seconde concernant la réalisation de l'enquête elle-même et sa diffusion (déclaration n°1210067 sur l' « enquête de cheminement tous niveaux » ; récépissé du 19 janvier 2007).

### 1.4.2. Le questionnaire allégé

Le questionnaire allégé comprend la même partie filtre que le questionnaire complet.

Les parties suivantes sont présentes sous forme simplifiée : parcours scolaire, calendrier professionnel, caractéristiques individuelles et environnement familial.

Les parties suivantes ne sont pas abordées : séjours à l'étranger, expériences de travail en cours d'études, perspectives professionnelles, mode de cohabitation.

Concernant le calendrier mensuel d'activité, l'ensemble des séquences sont mentionnées mais seuls le premier emploi et la situation à la date de l'enquête sont décrits, de manière plus succincte que dans le questionnaire complet (la formation, l'autonomie dans l'activité professionnelle, le bénéfice d'une promotion interne, l'opinion sur sa situation d'emploi ne sont pas abordés).

### 1.4.3. Zoom sur le calendrier professionnel

Le calendrier professionnel est le cœur du questionnaire. Il retrace mois par mois le parcours de l'enquêté entre sa date de fin d'études et la date de l'enquête et pilote l'ouverture des différents modules qui permettront un descriptif détaillé des séquences qui y sont repérées.

Figure 4. Le calendrier professionnel

	2004												2005												2006												2007					
	Jan	Fev	Mar	Avr	Mai	Jun	Jul	Aou	Sep	Oct	Nov	Dec	Jan	Fev	Mar	Avr	Mai	Jun	Jul	Aou	Sep	Oct	Nov	Dec	Jan	Fev	Mar	Avr	Mai	Jun	Jul	Aou	Sep	Oct	Nov	Dec	Jan	Fev	Mar	Avr	Mai	Jun
INTERIM																																										
ENTREPRISE																																										
RECHERCHE D'EMPLOI																																										
REPRISE D'ETUDES																																										
FORMATION																																										
AUTRES SITUATIONS																																										

Note : exemple pour une fin d'étude en juin 2004.

Le calendrier permet de distinguer six situations exclusives l'une de l'autre : deux situations d'emploi : intérim ou autre (« en entreprise ») ; une situation de recherche d'emploi ; une situation de reprise d'études ; une situation de formation<sup>2</sup> et les autres cas<sup>3</sup>. Par construction, une seule situation peut être déclarée pour un mois donné et toute séquence dure un nombre entier de mois. Deux séquences successives ne peuvent pas correspondre à une même situation (par exemple, il n'est pas possible de décrire deux séquences d'intérim

<sup>2</sup> Les situations de contrat d'apprentissage, de contrat de qualification ou d'élève fonctionnaire sont considérées comme des périodes d'emploi « en entreprise » et traitées comme telles.

<sup>3</sup> En particulier, si l'enquêté déclare une période d'emploi sur la période d'été suivant sa sortie de formation, il est demandé s'il s'agit d'un emploi de vacances. Dans ce cas, la séquence est codée « autre situation ».

contiguës ou deux séquences de formation contiguës). Par exception, deux séquences en entreprises peuvent être contiguës en cas de changement d'établissement (mais pas en cas de changement de profession ou de contrat au sein du même établissement).

Lors de l'enquête, l'ensemble des séquences d'emploi ou de non emploi sont décrites de façon successive en commençant par celle qui débute à la date de sortie de formation initiale et en terminant par la séquence correspondant à la situation actuelle. Lorsque le calendrier professionnel est rempli pour l'intégralité de la période, c'est-à-dire de la date de sortie de formation initiale augmentée d'un mois à la date d'enquête, la succession des séquences est listée à l'enquêté pour être validée.

Après cette validation, des modules de description fine des séquences s'ouvrent par ordre chronologique. Pour le questionnaire complet, toutes les séquences sont concernées<sup>4</sup>. Pour le questionnaire allégé, seules la première séquence d'emploi et la séquence relative à la situation à la date de l'enquête sont concernées.

### **Encadré 3 – Les informations demandées sur les différentes séquences du calendrier mensuel d'activité**

*Pour les séquences d'emploi passées* : nom de l'entreprise, lieu de travail, appartenance au secteur public, activité et taille de l'entreprise ; contexte de l'embauche, raison de l'acceptation du poste ; nature du contrat de travail, profession, durée du travail, rémunération à l'embauche ; mêmes informations à la fin de période d'emploi dans l'entreprise ; formation au cours de la période d'emploi ; motif du départ.

*Pour les séquences de non emploi passées* (recherche d'emploi, formation, autres) : inscription aux Assedic, démarches de recherche d'emploi, formation.

*Pour les séquences de reprise d'études passées* : but (obtenir un diplôme, préparer un concours public) et rôle du service public de l'emploi dans la reprise.

*Pour les séquences d'emploi à la date d'enquête* : nom de l'entreprise, lieu de travail, appartenance au secteur public, activité et taille de l'entreprise ; contexte de l'embauche, raison de l'acceptation du poste ; nature du contrat de travail, profession, durée du travail, rémunération à l'embauche ; mêmes informations à la date d'enquête ; formation, autonomie dans l'activité professionnelle ; bénéfice d'une promotion interne ; opinion sur sa situation d'emploi (déclassement subjectif et souhait de mobilité).

*Pour les séquences de non emploi à la date d'enquête* (recherche d'emploi, formation, autres) : inscription aux Assedic, démarches de recherche d'emploi, formation ; revenus, dont bénéfices d'allocations.

*Pour les séquences de reprise d'études à la date d'enquête* : but (obtenir un diplôme, préparer un concours public), rôle du service public de l'emploi dans la reprise, diplôme ou type de concours préparé.

---

<sup>4</sup> Dans le cas où plusieurs missions d'intérim se succèdent pendant plusieurs mois consécutifs, le calendrier ne comporte qu'une seule séquence « intérim » de plusieurs mois. Une seule mission sera décrite : la première mission dans le cas où celle-ci correspond au premier emploi après la fin d'étude ; la dernière mission dans le cas où celle-ci correspond à la situation actuelle ; sinon, c'est la mission la plus longue qui sera décrite.

#### **1.4.4. Les extensions de questionnement**

- Module « Post-initiaux des formations du domaine Santé/Social »

Le module est associé au questionnaire complet. Il concerne uniquement les sortants de l'extension d'échantillon Santé/Social qui ne sont pas primo sortants. Pour eux, le module se substitue aux questions sur le parcours scolaire et les expériences de travail en cours d'études.

Les questions abordent l'existence d'une formation antérieure dans le domaine, la situation d'activité avant la dernière formation, les objectifs de cette dernière, l'origine de son financement, le plus haut diplôme obtenu, le baccalauréat obtenu.

- Module « thèse »

Le module est associé au questionnaire complet. Il concerne tous les sortants de thèse ayant réussi leur doctorat interrogés avec le questionnaire complet (exceptés ceux du domaine Santé). Lors de la passation du questionnaire, il est posé après les questions sur le parcours scolaire et les séjours à l'étranger dans le cadre de la scolarité.

Les questions du module abordent le projet professionnel au moment de la soutenance de thèse, le cadre institutionnel de la thèse (nature de la rémunération, nature du laboratoire), la publication de travaux avant la soutenance, la situation post doctorale (stages post doctorat, qualification au CNU).

- Module « Apprentissage »

Le module est associé au questionnaire complet. Il concerne tous les sortants d'une formation dispensée par voie d'apprentissage. Lors de la passation du questionnaire, il est posé dans la partie qui aborde le parcours scolaire.

Les questions du module abordent l'activité et la taille de l'entreprise ayant accueilli le jeune en apprentissage, comment l'entreprise a été trouvée, la période du contrat d'apprentissage, les conditions de travail (horaires et jours), le rôle de l'emploi en entreprise dans la formation du jeune, l'embauche ultérieure ou non dans cette entreprise.

- Module « Santé Handicap »

Le module est associé au questionnaire complet. Il concerne tous les répondants, qu'ils soient dans une extension d'échantillon ou non. Il est passé en fin de questionnaire, avant la partie consacrée aux discriminations ressenties.

Cinq questions abordent successivement l'existence d'une gêne dans le travail en raison d'un handicap ou d'un problème de santé, le dépôt d'un dossier auprès de la maison départementale des personnes handicapées, de la COTOREP ou de la CDES, la décision rendue après ce dépôt, l'état de santé général déclaré.

- Module « Post-Initiaux des formations du sport et de l'animation »

Le module est associé au questionnaire allégé. Il concerne uniquement les diplômés de l'extension d'échantillon sur les formations du sport et de l'animation qui ne sont pas primo sortants. Pour eux, le module se substitue aux questions sur le parcours scolaire.

Les questions abordent l'existence d'une formation antérieure dans le domaine du sport ou de l'animation, la situation d'activité avant la dernière formation, les objectifs de cette formation, l'origine de son financement, le plus haut diplôme obtenu, le baccalauréat obtenu, la détention de diplômes et brevets dans le domaine du sport et de l'animation.

- Module « Multi-activité »

Le module est associé au questionnaire allégé. Il concerne uniquement les diplômés de l'extension d'échantillon sur les formations du sport et de l'animation, primo sortants ou non. Lors de la passation du questionnaire, il est posé à l'issue de la description du premier emploi et de l'emploi actuel : le module aborde l'existence d'une activité secondaire parallèle dans le domaine du sport ou de l'animation.

### 1.5. Le calendrier de l'enquête

La réalisation de l'enquête s'est étendue sur quatre années. Les bases de sortants de formation initiale ont été collectées dès leur disponibilité, au cours de l'année 2005. Le questionnaire a été élaboré au cours de l'année 2006 et du premier trimestre 2007. Les extensions ont été définies sur la même période. L'enquête a été réalisée par téléphone lors du 2<sup>nd</sup> trimestre 2007, la phase de pré qualification commençant dès le 1<sup>er</sup> trimestre 2007. La figure 5 reprend de façon plus détaillée la chronologie des principales étapes de l'enquête « Génération 2004. »

**Figure 5. Calendrier de l'opération**

	2005			2006				2007				2008			
	2nd trim.	3ème trim.	4ème trim.	1er trim.	2nd trim.	3ème trim.	4ème trim.	1er trim.	2nd trim.	3ème trim.	4ème trim.	1er trim.	2nd trim.	3ème trim.	4ème trim.
<b>Opérations Enquête Génération 2004</b>															
Collecte des bases établissements	x		x												
Conception et test papier des questionnaires (complet et allégé)					x	x	x	x							
Elaboration des conventions d'extensions						x	x	x							
Conception des Catis pour les questionnaires et tests par le Céreq								x							
Tests des questionnaires Catis chez H2A								x							
Tirage de l'échantillon								x							
Recherche des coordonnées téléphoniques des individus								x							
Terrain de pré-qualification								x	x						
Terrain d'enquête									x						
Constitution des bases, apurements, codifications, pondération, rédaction de la documentation										x	x	x	x		
Livraison des bases aux extensions													x		

#### **Encadré 4 – Le « test papier » des questionnaires**

Afin de vérifier la bonne compréhension des questions par les enquêtés, des tests des questionnaires complet et allégé ont été réalisés avant la finalisation des questionnaires et leur mise sous Cati. Ils ont également permis une première estimation de la durée : une durée trop longue peut conduire les enquêtés à mettre fin au questionnement avant la fin de passation du questionnaire.

Le questionnaire complet a été testé auprès d'individus appartenant au champ de l'enquête (voir 1.2), entre le 25 septembre et le 10 octobre 2006, en fin de journée pour pouvoir joindre des personnes en emploi, entre 17h à 20h30. Quarante-six questionnaires exploitables ont été collectés.

Le questionnaire allégé a été testé dans son intégralité entre le 11 et le 13 octobre 2006, sur la même tranche horaire. L'objectif était plus particulièrement de tester l'enchaînement des questions et le minutage, ce questionnement court étant nouveau. Au final, 39 questionnaires exploitables ont été collectés.

Aucune difficulté majeure n'a été mise en évidence lors des tests papiers. Ils ont néanmoins donné lieu à quelques ajustements :

- concernant les « Jobs » de vacances, la question sur la fréquence (« souvent » / « parfois ») a été précisée avec l'introduction d'un seuil (trois fois par an) ;
- nécessité d'introduire la possibilité de réponses multiples à la question sur la raison pouvant conduire à refuser un emploi ;
- précision d'un nombre d'heures indicatif pour les questions sur le taux du temps partiel, pour tenir compte du fait que la base horaire hebdomadaire est variable.

## 2. LA CONSTITUTION DE LA BASE DE SONDAGE

---

Il n'existe pas de base d'élèves centralisée couvrant l'ensemble des formations initiales dispensées en France métropolitaine. Il faut donc la constituer spécifiquement pour l'enquête, à partir de différentes sources, dans le cadre d'une collecte *ad hoc*.

La base ainsi constituée présente un défaut de couverture dû aux établissements qui n'ont pas fourni de bases d'élèves lors de la constitution de la base de sondage.

Par ailleurs, les données récupérées sont souvent des listes de sortants d'un établissement<sup>5</sup> : il s'agit de personnes inscrites en 2003-2004 dans l'établissement et non réinscrites en 2004-2005 dans le même établissement. Elles peuvent néanmoins s'être réinscrites ensuite ailleurs et sont alors hors du champ de l'enquête. La base comporte aussi des doublons : les jeunes inscrits dans plusieurs établissements ou dans plusieurs filières. Il n'est possible de traiter que partiellement ces deux aspects lors de la constitution de la base de sondage.

### 2.1. La collecte auprès des établissements

Pour chaque individu scolarisé pendant l'année scolaire 2003-2004 (présupposé sortant), les informations suivantes devaient être récupérées : nom et prénom, adresse et numéro de téléphone, date de naissance, sexe, diplôme et spécialité de formation, obtention ou non du diplôme préparé.

Des informations relatives à l'établissement de formation étaient également collectées. Lorsque le numéro d'identifiant national d'étudiant (INE) était disponible (informations issues des rectorats, des universités et des CFA), il était récupéré pour les traitements d'apurement de la base (repérage des poursuites d'études et des doublons). Dans ce cas, lorsque l'établissement disposait du fichier d'inscrits en 2004-2005, ce fichier était également récupéré.

#### 2.1.1. Collecte des bases rectores

Les rectorats disposent des listes d'inscrits dans les établissements de leur ressort : collèges, lycées et BTS dépendant du ministère en charge de l'éducation nationale (y compris les établissements privés sous contrat), grâce au système informatique *Scolarité*.

Le Céreq a réalisé lui-même la collecte des informations souhaitées auprès des rectorats, initialisée par un courrier appuyé par la DEPP, et selon un dessin d'extraction également défini avec la DEPP.

La collecte a eu lieu au second trimestre 2005. Elle a permis de récupérer pour chaque académie l'ensemble des inscrits au cours de l'année scolaire 2003-2004, ainsi que celui des inscrits en 2004-2005.

---

<sup>5</sup> Certains établissements d'enseignement supérieur comme les écoles d'ingénieurs ou de commerce n'ont transmis que les informations concernant les diplômés. La concordance avec les sortants est élevée car dans ces établissements, le taux d'échec ou d'abandon en cours d'étude est très faible.

La couverture des établissements est ainsi exhaustive sur le champ rectoral. Au total, les établissements couverts par les rectorats représentent 82% des établissements de la base de sondage finale et concernent, d'après les résultats de l'enquête, 42% des sortants<sup>6</sup>.

### *2.1.2. Collecte auprès des autres établissements*

Selon la banque centrale des établissements (BCE)<sup>7</sup>, gérée par le ministère en charge de l'éducation nationale, 3 724 établissements n'entraient pas dans le champ des bases disponibles dans les rectorats : universités, « grandes écoles », lycées agricoles, centres de formation des apprentis,....

Le Céreq a directement collecté les données pour un petit nombre d'établissements dépendant de certains ministères et a confié à un prestataire les contacts sur le reste de la collecte, qui s'est déroulée de novembre à décembre 2005, en plusieurs phases :

- *Phase 1* : envoi d'un courrier signé par le directeur du Céreq sollicitant la mise à disposition des données individuelles nécessaires à la base de sondage ; obtention de l'accord de principe du responsable de l'établissement et désignation d'un correspondant technique.
- *Phase 2* : envoi d'un courrier au correspondant technique pour l'aider à constituer le fichier souhaité, échanges et fixation d'une date de transmission.
- *Phase 3* : relance de l'interlocuteur en cas de non réception du fichier à la date convenue. Cinq vagues de relances maximum ont été effectuées.

Hormis les contacts initiaux par courrier, les contacts avec les établissements ont été téléphoniques et assistés par un système Cati. Afin de diminuer les taux de refus des établissements, une dizaine de ministères de tutelle avaient été préalablement sensibilisés à l'enquête.

Les fichiers étaient réceptionnés directement par le Céreq qui a systématiquement vérifié la lisibilité et la conformité des données (année de référence correcte, disponibilité de l'ensemble des informations attendues).

Après exclusion des établissements hors champs (204) et des faux numéros (224), 3 296 adresses se sont révélées exploitables et 2 354 établissements ont transmis leurs informations, toujours directement au Céreq.

---

<sup>6</sup> Par ailleurs, les universités représentent 0,6% des établissements de la base de sondage finale et 25% des sortants ; les IUFM, respectivement 0,2% et 3% et les autres établissements, 18% et 30%.

<sup>7</sup> La BCE est le répertoire national des établissements assurant une activité de formation initiale générale, technique ou professionnelle, de la maternelle à l'enseignement supérieur, qu'ils soient publics ou privés, sous tutelle ou non du ministère de l'éducation nationale et de la recherche. Elle concerne également les structures d'administration du système éducatif public ainsi que certaines parties d'établissements qui ont besoin d'être identifiées pour la gestion du système éducatif. La BCE a pour rôle d'alimenter toutes les applications informatiques nationales de gestion administrative, financière, statistique ou documentaire du ministère de l'éducation nationale et de la recherche. Elle est mise à jour à partir des bases rectorales d'établissements, en temps réel. Source : <http://www.infocentre.education.fr>.

**Figure 6. Détail de la réception des fichiers par type d'établissement  
(hors bases collectées via les rectorats)**

<b>Bilan final de la collecte auprès des établissements</b>	<b>Nombre total d'adresses d'établissements</b>	<b>Nombre d'adresses d'établissements exploitables</b>	<b>Nombre de fichiers d'établissements reçus</b>	<b>Soit en % du nombre d'adresses d'établissements exploitables</b>
Ecoles de l'enseignement supérieur	1390	1201	753	62,7
Universités	78	78	69	88,5
Ecoles, Collèges, Lycées secondaires absents des fichiers Rectorat	1234	1090	793	72,8
Institut Universitaire de Formation des Maitres (IUFM)	26	26	22	84,6
Centres de formation des apprentis (CFA)	980	885	709	80,1
Ecoles de la fonction publique	16	16	8	50,0
<b>Ensemble</b>	<b>3724</b>	<b>3296</b>	<b>2354</b>	<b>71,4</b>

## 2.2. L'apurement de la base de sondage

Une fois constituée, la base a été apurée pour supprimer, quand cela était possible, les individus hors champ et les doublons.

Pour les bases transmises par les rectorats, la disponibilité d'un identifiant commun de gestion (l'Identifiant National Etudiant ou INE) et des fichiers des inscrits en 2003-2004 et en 2004-2005 a permis de repérer de manière précise les poursuites d'études internes à ce champ en comparant les fichiers des inscrits en 2003-2004 et les fichiers des inscrits en 2004-2005. La poursuite d'études vers des établissements en dehors de ce champ, comme le passage du champ des rectorats ou des centres de formation des apprentis vers l'université ne sont pas toujours repérables à partir de l'INE. Une partie des poursuites d'études ont toutefois été repérées à partir des nom, prénom et date de naissance (mois, année) entre les fichiers des inscrits en 2003-2004 et les fichiers des inscrits en 2004-2005.

Il reste donc de nombreux individus hors champ dans la base de sondage finale. Les individus poursuivants qui n'ont pas pu être repérés lors de cette étape sont mis hors champ lors de la passation du questionnaire filtre. Pour l'enquête « Génération 2004 », le repérage de personnes hors champ lors de la collecte a concerné 52 000 personnes, soit 44% des personnes contactées ayant accepté de répondre.

Outre les poursuites d'études, des doublons parmi les inscrits de 2003-2004 sont également repérés. Ils le sont à partir des nom, prénom et date de naissance (mois, année). Le dé-doublonnage est effectué sur la base complète, tous types d'établissements confondus.

Les individus âgés de plus de 35 ans en 2004 ont également été supprimés, sauf pour les extensions de champ prévues pour les formations du sport et de l'animation et du domaine santé/social de manière à conserver les individus post-initiaux.

Contrairement aux enquêtes « Génération 1998 » et « Génération 2001 », les étudiants étrangers ont été conservés (pour les autres niveaux, les étrangers sont aussi conservés).

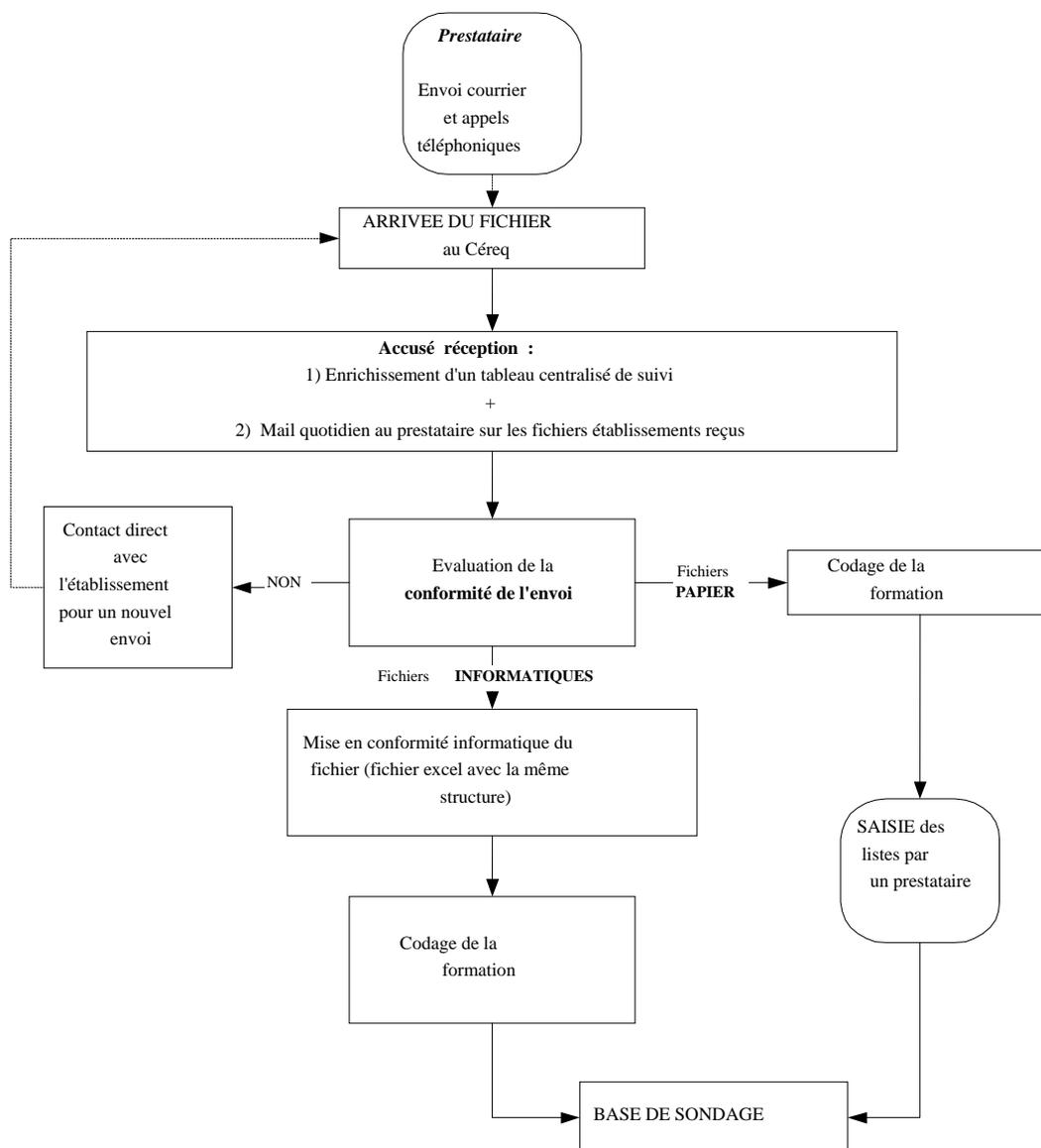
Au final, la base de sondage de l'enquête « Génération 2004 » comportait 1 268 532 individus répartis comme présenté dans la figure 7.

**Figure 7. Effectifs dans la base de sondage par type d'établissement**

Type d'Etablissement	Effectifs dans la Base
Rectorats	674 405
Université	324 959
CFA	123 957
Lycées Agricoles	35 557
IUFM	19 367
Ecoles de la Santé	17 395
Ecoles de Commerce	17 077
Ecoles d'Ingénieur	14 880
Directions Régionales Jeunesse et Sport (DRJS)	12 900
Ecoles du Ministère de la Culture	8 148
Ecoles des professions sociales	6 345
Facultés privées	3 793
Divers	3 403
Ecoles secteur services	1 776
Centres d'Enseignement privé	1 221
Instituts d'Etudes Politiques (IEP)	1 205
Ecoles formation agricole	823
Ecoles secteur industrie	448
Ecoles Normales supérieures	386
Ecoles de l'Administration publique	250
Ecoles de la Direction Générale de l'Aviation Civile	237
Ensemble	1 268 532

La figure 8 présentée ci-dessous résume l'ensemble des opérations mises en œuvre pour collecter les fichiers établissements et intégrer les informations dans la base de sondage.

**Figure 8. Schéma récapitulatif de la collecte des fichiers établissements  
(hors fichiers transmis par les rectorats)**



### 3. LA CONSTITUTION DE L'ÉCHANTILLON

---

#### 3.1. Le cadre général du plan de sondage

Le plan de sondage de l'enquête « Génération 2004 » a été construit de façon à obtenir environ 70 000 questionnaires, 40 000 questionnaires complets et 30 000 questionnaires allégés, répartis de façon à satisfaire les différentes demandes d'extension et selon un tirage équilibré par grands niveaux de formation et régions.

La base de sondage a été divisée en huit ensembles qui correspondent à 8 champs du plan de sondage. Cette stratification est construite à partir du type d'établissements, les apprentis étant mis à part puisqu'une estimation spécifique de la couverture est disponible pour eux : « rectorats », « universités », « formations agricoles », « Jeunesses et Sport » (établissements concernés par l'extension Sport), « écoles de la Santé et du Social » (établissements concernés par l'extension Santé Social), « IUFM », « écoles du supérieur » et « apprentis ». Cette dernière strate comprend tous les apprentis, qu'ils soient sortants de CFA ou non. Les autres regroupements s'entendent donc hors apprentis. Par exemple, le champ « Université » regroupe les sortants des universités hors apprentis.

Le tirage de l'échantillon a été effectué séparément pour chacun de ces huit champs.

Pour la première fois, un tirage à deux degrés a été mis en place afin de permettre, en aval, des analyses multi-niveaux, dans le but de voir l'existence possible d'un effet lié à l'établissement de formation dans la vitesse d'insertion des jeunes sortis de ces établissements. Dans les champs pour lesquels cela était possible, des établissements ont d'abord été tirés puis des sortants de ces établissements. Sont concernés les collèges et lycées couverts par le champ Rectorat et les CFA. Dans les six autres champs, il n'y a pas eu de tirage préalable d'établissements. Au final le tirage à deux degrés concerne 149 000 individus sur les 403 000 de l'échantillon (soit environ 37 %).

*In fine*, le tirage est donc stratifié, à deux degrés sur les champs qui le permettent, et à probabilités inégales. La figure 9 présente ses principales caractéristiques.

Le tirage a été réalisé à l'aide de la macro procédure « Cube » développée par l'Insee. Cet algorithme d'échantillonnage permet d'améliorer la précision des estimateurs associés aux variables d'intérêt de l'enquête<sup>8</sup>. Pour cela, Cube permet de tirer de manière aléatoire un échantillon équilibré sur un ensemble de totaux connus à partir d'informations auxiliaires disponibles dans la base de sondage : la méthode consiste à choisir un échantillon tel que les estimateurs d'Horvitz-Thompson des totaux des variables choisies pour l'équilibrage coïncident avec les vrais totaux. Cette procédure s'applique sur des bases de sondage munies d'informations auxiliaires, qualitatives ou quantitatives, connues au niveau individuel. Des probabilités d'inclusion individuelles destinées à surreprésenter les individus inclus dans des champs d'extensions ont ainsi pu être intégrées dans la procédure de tirage.

Pour l'enquête « Génération », l'équilibrage initial, avant le calcul des coefficients d'extension d'échantillon, a été fait de façon à respecter la répartition des effectifs selon les niveaux de sortie, affinées pour les formations professionnelles du champ « Rectorat » et « CFA » pour respecter la répartition entre formations industrielles et formations tertiaires.

---

<sup>8</sup> Pour plus d'informations, se reporter à la *Documentation de l'utilisateur de la Macro CUBE* rédigée par S. Rousseau et F. Tardieu (2004).

**Figure 9. Caractéristiques des tirages à l'intérieur de chaque champ**

<p style="text-align: center;"><b>Rectorats</b></p> <hr/> <ul style="list-style-type: none"> <li>• Trois tirages :               <ul style="list-style-type: none"> <li>- collèges</li> <li>- lycées professionnels</li> <li>- lycées généraux et technologiques.</li> </ul> </li> <li>• Tirages 2 niveaux</li> <li>• Tirages équilibrés sur :               <ul style="list-style-type: none"> <li>- la spécialité (industrielle ou tertiaire)</li> <li>- le niveau</li> <li>- la région de l'établissement.</li> </ul> </li> </ul>	<p style="text-align: center;"><b>Universités</b></p> <hr/> <ul style="list-style-type: none"> <li>• Trois tirages :               <ul style="list-style-type: none"> <li>- licence 1 à licence 3 (L1 à L3)</li> <li>- maîtrise 1 et 2 (M1, M2)</li> <li>- doctorat.</li> </ul> </li> <li>• Tirages équilibrés sur :               <ul style="list-style-type: none"> <li>- la spécialité (disciplines SISE)</li> <li>- le niveau</li> <li>- la région de l'établissement.</li> </ul> </li> </ul>	<p style="text-align: center;"><b>Apprentis (CFA et hors CFA)</b></p> <hr/> <ul style="list-style-type: none"> <li>• Tirage 2 niveaux (sauf Corse, Limousin, Lorraine, Poitou-Charentes)</li> <li>• Tirage équilibré sur :               <ul style="list-style-type: none"> <li>- la spécialité (industrielle ou tertiaire)</li> <li>- le niveau</li> <li>- la région de l'établissement.</li> </ul> </li> </ul>
<p style="text-align: center;"><b>Formations agricoles</b></p> <hr/> <ul style="list-style-type: none"> <li>• Tirage équilibré sur la strate de formation (A10 : bac professionnel, A20 : BTA, A30 : bac technologique, A40 : CAPA BPA BEPA, A50 : BTSA)</li> </ul>	<p style="text-align: center;"><b>DRJS</b></p> <hr/> <ul style="list-style-type: none"> <li>• Tirage équilibré sur la strate de formation (J01 : brevets d'état, J02 : BPJEPS, J03 : BAPAAT, J05 : BEATEP)</li> </ul>	<p style="text-align: center;"><b>Ecoles de la santé et du social</b></p> <hr/> <ul style="list-style-type: none"> <li>• Tirage équilibré sur la strate de formation (s10, s20, s30, s40, s50)</li> </ul>
<p style="text-align: center;"><b>IUFM</b></p> <hr/> <ul style="list-style-type: none"> <li>• Tirages équilibrés sur la strate de formation (I10 : sortants de 1<sup>ère</sup> année, I20 : passage en 2<sup>ème</sup> année)</li> </ul>		<p style="text-align: center;"><b>Ecoles du Supérieur</b></p> <hr/> <ul style="list-style-type: none"> <li>• Quatre tirages :               <ul style="list-style-type: none"> <li>- écoles d'Ingénieurs</li> <li>- écoles de Commerce</li> <li>- écoles de la Culture</li> <li>- autres écoles du Supérieur</li> </ul> </li> <li>• Tirages équilibrés sur la strate de formation</li> </ul>

Les poids de tirage ont été calculés en tenant compte d'estimations sur le taux de couverture de la base de sondage et du taux de réponse constaté à l'enquête « Génération 1998 » (ce dernier taux incluant en réalité trois aspects qui ne sont pas distingués : la probabilité de contact, l'acceptation et le fait de s'avérer être bien dans le champ).

Le taux de tirage se décompose alors ainsi :  $TXI = P1 * P2 * P3$ , avec :

- $P1$ , le coefficient de dilatation lié à la couverture
- $P2$ , le coefficient de dilatation lié à la non-réponse attendue
- $P3$ , le coefficient d'échantillonnage.

Les parties suivantes détaillent les principes d'estimation de ces trois composantes. Il convient ici de souligner que si l'enquête « Génération 2004 » comporte un échantillon national et des extensions d'échantillon, le tirage de l'échantillon national et celui des extensions d'échantillon ne sont toutefois pas disjoints. L'échantillon national et ses extensions sont tirés simultanément. A cet effet, le coefficient d'échantillonnage au sein de chaque strate ( $P3$ ) est adapté de façon à assurer les sur représentations demandées par les extensions, compte tenu du taux de couverture de la base ( $P1$ ) et du taux de non réponse prévisible des individus ( $P2$ ) déjà estimé.

### **3.2. L'estimation du taux de couverture des individus entrant dans le champ ( $P1$ )**

L'estimation de la couverture de la base de sondage, qui renvoie au degré d'exhaustivité de la collecte des fichiers d'établissement, a été effectuée sur chacun des huit champs de tirage en s'appuyant sur diverses sources administratives. Sauf mention contraire, les taux de couverture présentés portent sur les effectifs couverts et non sur le nombre d'établissements couverts.

L'annexe 1 présente les valeurs moyennes de  $P1$  par strate agrégée (variable Strateg) au sein des huit champs.

Pour le champ « Universités » (*hors apprentis*) :

Un total de 69 universités sur 78 ont transmis leurs fichiers de sortants.

Les effectifs de la base considérés comme sortants après le repérage des doublons sur le champ université ont été comparés à ceux fournis par le système d'information de suivi des étudiants de la DEPP (système SISE).

Au niveau national, la couverture du champ universitaire est plutôt bonne, avec un taux de couverture de 84,8 %. Il existe toutefois des disparités régionales du fait de la difficulté rencontrée pour récupérer des fichiers de sortants exploitables auprès de certaines universités.

**Figure 10. Estimation de la couverture des universités par niveau (hors apprentis)**

Niveau	Effectifs dans la base de sondage	Taux de couverture estimé
L1 en IUT	8 562	69,3
Autres L1	102 410	76
L2 en IUT	19 483	92,6
Autres L2	30 185	86,3
L3 professionnel	11 429	95,9
Autres L3	60 871	89,1
M1	68 306	89,1
M2	57 121	88,4
Docteurs	12 213	81,3
Autres Sortants	9 504	81,9
<b>Ensemble</b>	<b>380084</b>	<b>84,8</b>

Le coefficient de dilatation lié à la couverture ( $P1$ ) a été estimé sur un croisement région - niveau- discipline.

Pour le champ « Apprentis » (CFA et hors CFA) :

80 % des établissements contactés ont transmis leurs fichiers d'inscrits. En termes d'effectifs, le coefficient de couverture dans la base de sondage est de 71,3 % pour le champ des apprentis. Il a été estimé à partir de l' « enquête n°51 sur les centres de formation d'apprentis », enquête annuelle pilotée par la DEPP du ministère de l'Éducation nationale sur l'ensemble des apprentis (CFA et hors CFA)<sup>9</sup>.

La correction de la couverture ( $P1$ ) sur le champ des apprentis a été estimée sur un croisement région – niveau - spécialité (industrielle, tertiaire) ainsi que par spécialité fine, selon la nomenclature des spécialités de formation (NSF<sup>10</sup>).

Pour le champ des formations agricoles (hors apprentis) :

En termes d'effectifs, le taux de couverture moyen est de 66 % lorsqu'on rapproche les effectifs estimés de ceux fournis par la base Regards sur les flux de l'enseignement technique et professionnel (Reflét)<sup>11</sup>. La couverture est un peu plus faible qu'ailleurs en Ile-de-France, Franche-Comté et Haute-Normandie.

<sup>9</sup> L'enquête fournit, sur différents échelons géographiques, des statistiques sur les effectifs par sexe, par âge, par niveau et spécialité de formation, par secteur d'activité de l'employeur (code APE). Les données sont collectées de façon agrégées.

<sup>10</sup> Voir *infra*, partie 5.2.7.

<sup>11</sup> La base Reflét, gérée par le Céreq, répertorie les diplômes de l'enseignement technique et professionnel délivrés par les ministères de l'Éducation nationale, de l'Agriculture et de la pêche, et les directions générales de l'Action sociale et de la Santé. Elle contient plus de 2 600 diplômes (hors licences professionnelles) pour lesquels elle propose notamment des statistiques sur les effectifs en dernière année de formation, ainsi que les résultats aux examens.

La comparaison des effectifs de la base de sondage avec ceux figurant dans la base Reflet n'est toutefois possible que sur une partie du champ. Le taux de couverture (*PI*) a donc été calculé par région et strate de formation sur les effectifs communs aux deux bases.

Pour le champ des formations « Jeunesse et Sport » (*hors apprentis*) :

En l'absence d'informations spécifiques, le champ a été défini à partir d'une liste de formations transmises par le ministère en charge des sports (à partir d'informations issues des directions régionales de la jeunesse et des sports). Toutes les formations figurant sur la liste transmise par le ministère demandeur de l'extension ayant été collectées, le champ des diplômés des formations du ministère des Sports a été considéré comme exhaustif. Le coefficient de redressement pour défaut de couverture (*PI*) est donc fixé à 1.

Pour le champ des formations du domaine de la Santé et du Social (*hors apprentis*) :

Le champ a été défini à partir d'une liste d'établissements transmise par la DREES. Au total, on comptabilisait un peu plus de 24 000 diplômés des écoles de la santé et du social en 2004 dans la base de sondage. Selon le fichier national des établissements sanitaires et sociaux (FINESS)<sup>12</sup>, ils étaient 53 200, soit une couverture moyenne de 45,3 % pour la base de sondage en termes d'effectifs. Cette faible couverture résulte essentiellement de la non-réponse totale ou partielle des établissements concernés.

Pour le champ des IUFM (*hors apprentis*) :

Vingt-deux établissements sur vingt-six ont transmis des fichiers de sortants. En termes d'effectifs, on dénombrait 19 362 sortants d'IUFM non apprentis dans la base. Selon les informations transmises par la DEPP, le taux de couverture atteint environ 52,6 % avec des différences assez fortes selon les régions. La correction de la couverture a été appliquée, dans la mesure du possible, par région.

Pour le champ des écoles du supérieur :

Un total de 63 % des écoles du supérieur contactées ont transmis des fichiers de sortants. Le taux de couverture estimé varie selon le type d'école dans les proportions suivantes :

- écoles de commerce (*hors apprentis*) : 16 444 sortants collectés sur 18 289 estimés soit une couverture de 89,9 % ;
- écoles d'ingénieurs (*hors apprentis*) : 14 652 sortants collectés sur 17 631 estimés soit une couverture de 83,1 % ;

---

<sup>12</sup> Le fichier FINESS est un fichier géré par le ministère en charge de la santé et des affaires sociales pour répondre d'abord à des besoins de gestion, puisqu'il permet l'immatriculation par les DDASS et les DRASS des établissements et entités juridiques porteurs d'une autorisation ou d'un agrément dans les trois domaines : le sanitaire et médico-social, le social, l'enseignement des personnels en santé et en social.

- écoles de la Culture (*hors apprentis*) : les sortants non apprentis des écoles de la Culture sont au nombre de 7 864 dans la base sur les 13 364 sortants estimés soit une couverture de 58,8 %.
- divers Ecoles du Supérieur (*hors apprentis*) : en l'absence d'éléments suffisants pour calculer un taux de couverture sur les autres champs du supérieur, nous n'avons pas déformé notre tirage ( $P1=1$ ).

La figure 11 fournit le taux couverture moyen estimé pour chaque champ de tirage en termes d'effectifs.

**Figure 11. Estimation de la couverture des effectifs par champ de tirage**

Champs de tirage	Taux de Couverture estimé
Fichiers des rectorats	100,0 %
Universités	84,8 %
CFA	71,3 %
Lycées Agricoles	66,3 %
IUFM	52,6 %
Ecoles de la Santé et du Social	45,3 %
<b>Ecoles du Supérieur</b>	
Ecoles de Commerce	89,9 %
Ecoles d'Ingénieur	83,1 %
Jeunesse et Sport	100,0 %
Ecoles de la Culture	58,8 %
Divers Ecoles du Supérieur	100,0 %

### 3.3. L'estimation du taux de non réponse prévisible des individus ( $P2$ )

Le coefficient de dilatation lié à la non-réponse a été estimé sur les croisements champ, strate de formation et région, le degré de finesse de la variable de strate de formation utilisée dépendant au cas par cas des effectifs présents dans la strate. Ils ont été calculés à partir des résultats de l'enquête « Génération 1998 à trois ans » : le coefficient est égal au nombre de questionnaires mis en production pour obtenir un répondant dans le champ. Il regroupe donc plusieurs phénomènes : la probabilité d'obtenir le bon contact, le taux de non-réponse à l'enquête (dont les refus) et la probabilité que la personne contactée et acceptant de répondre se révèle hors-champ.

L'enquête « Génération 1998 à trois ans » a été préférée à l'enquête « Génération 2001 à trois ans », plus récente, parce que la première disposait d'un échantillon bien plus important, ce qui permettait de calculer les coefficients sur des croisements plus fins.

L'annexe 1 présente les valeurs moyennes de  $P2$  par strate agrégée (variable Strateg) au sein des huit champs.

### 3.4. L'estimation du coefficient d'échantillonnage (P3)

Les coefficients d'échantillonnage  $P3$  sont estimés par étape en fonction du nombre de répondants visés par le Céreq et des sur représentations attendues pour les extensions d'échantillon. Ils sont estimés en tenant compte du taux de couverture de la base ( $P1$ ) et du taux de non réponse prévisible des individus ( $P2$ ).

#### 3.4.1. Première étape : le tirage à blanc d'un échantillon hors extension

Pour commencer, un premier tirage a été effectué en fixant un coefficient d'échantillonnage  $P3$  uniforme, égal à  $1/25^e$ , pour l'ensemble de la base. Chaque individu  $i$  s'est donc vu attribuer une probabilité d'inclusion égale à  $Taux1(i) = P1(i) \times P2(i) / 25$  dépendant de ses caractéristiques (champ, région de l'établissement de formation, strate de formation). S'il n'y avait pas de problème de couverture ni de taux de réponse prévisible différencié, il s'agirait donc d'un tirage à probabilité égale, avec un taux calibré pour atteindre environ 30 000 répondants.

Le tirage a ensuite été effectué selon les mêmes principes que ceux retenus pour l'échantillon final : des tirages région par région et, au sein de chaque région, champ par champ, à l'aide de la macro procédure « Cube », utilisant comme contraintes d'équilibrage la répartition des effectifs selon les modalités de la variable « Strateg » et, pour les formations professionnelles du champ « Rectorat » et « CFA », la répartition entre formations industrielles et formations tertiaires.

Ce premier tirage ne conduit pas à sélectionner des individus à enquêter mais permet d'obtenir une estimation des effectifs de répondants prévisibles dans le champ des différentes extensions demandées en l'absence d'extension d'échantillon.

#### 3.4.2. Deuxième étape : le calcul des coefficients de suppléments de tirage

Le premier tirage permet de calculer des coefficients de suppléments de tirage relatifs à chaque extension, par strate fine, chacun étant calculé ainsi :

$$Srs(S) = \text{nombre de questionnaires supplémentaires « souhaités » dans la strate } S \\ / \text{ nombre de questionnaires obtenus suite à la première étape de tirage}$$

Autant de coefficients de supplément de tirage sont calculés qu'il y a d'extensions nationales<sup>13</sup> :  $sr\_drees(S)$ ,  $sr\_anfa(S)$ ,  $sr\_js(S)$ ,  $sr\_forco(S)$ ,  $sr\_ifen(S)$ ,  $sr\_gse(S)$ ,  $sr\_depp(S)$ .

Le découpage régional servant à la constitution des strates, un seul coefficient  $sr\_reg(S)$  est calculé pour les extensions régionales.

Pour un individu présent dans la base de sondage au sein de la strate  $S$ , son poids de tirage devient :

---

<sup>13</sup> Excepté pour l'extension des formations de la Culture et pour l'extension des sortants de doctorat : ces deux extensions d'échantillon étant exhaustives sur leur champ, tous les individus concernés ont été échantillonnés.

$$Taux2(i) = Taux1(i) * (1 + sr\_reg(S) + sr\_drees(S) + sr\_anfa(S) + sr\_js(S) + sr\_forco(S) + sr\_ifen(S) + sr\_gse(S) + sr\_depp(S))$$

Pour certaines strates, l'application des coefficients de supplément de tirage liés aux extensions conduit à un taux de tirage supérieur à 1 ( $Taux2(i) > 1$ ). Autrement dit, il y a saturation de ces strates qui sont alors échantillonnées de façon exhaustive.

Le tirage final s'est donc fait à partir du poids de tirage  $Taux3(i) = \text{Min}[Taux2(i), 1]$ .

### 3.5. L'affectation des personnes échantillonnées au questionnaire « complet » ou au questionnaire « allégé »

Le tirage de l'échantillon s'effectue en une seule fois à partir des poids de tirage finaux sans distinguer *a priori* ceux qui relèvent de l'échantillon national de ceux qui relèvent de telle ou telle extension. Quand il y a lieu, les extensions de questionnement sont ainsi posées indifféremment à toutes les personnes enquêtées entrant dans le champ concerné.

Pour l'enquête « Génération 2004 à trois ans », la distinction faite entre un questionnaire complet et un questionnaire allégé nécessitait toutefois de répartir l'échantillon entre les personnes concernées par l'un et celles concernées par l'autre, l'objectif étant de poser le questionnaire complet aux personnes relevant de l'échantillon national du Céreq, ainsi qu'aux extensions d'échantillon suivantes : les sortants de doctorat, les sortants de formations sanitaires et sociales (pour la Drees) et les sortants des filières automobiles (pour l'Anfa).

Pour cela, les individus se sont vus affecter un poids de tirage correspondant à la probabilité d'être tiré au titre de l'échantillon national ou de l'extension de la Drees ou de l'Anfa, conditionnellement à la probabilité d'être tiré dans l'échantillon de l'enquête.

Ce calcul suppose d'établir une hiérarchie implicite dans l'ordre du tirage de l'échantillon national et des différents suppléments de tirage. Car, dans le cas des strates saturées, le nombre d'individus présents dans la base est inférieur à celui que l'on souhaiterait tirer au titre de l'échantillon national et des différents suppléments d'échantillon. Or, pour savoir s'ils sont concernés par le questionnaire complet ou le questionnaire allégé, il faut arbitrer pour dire s'ils sont considérés comme relevant de l'échantillon national, de l'extension de la Drees ou de l'Anfa, ou pas.

Le choix suivant a été fait : les personnes tirées sont d'abord considérées comme relevant de l'échantillon national, puis des extensions de la Drees et de l'Anfa, puis des autres.

Avec la même notation que précédemment, on pose :

$$sr\_comp = sr\_anfa + sr\_drees ;$$

$$sr\_all = sr\_gse + sr\_reg + sr\_depp + sr\_js + sr\_forco + sr\_ifen ;$$

Les cas de saturation sont alors traités comme suit :

*Si  $\text{taux1}(i) > 1$ ,  $\text{sr\_comp}^* = 0$  et  $\text{sr\_all}^* = 0$  ;*

*Sinon, si  $\text{taux1}(i) \times (1 + \text{sr\_comp}) > 1$ ,  $\text{sr\_comp}^* = (1/\text{Taux1}(i)) - 1$   
et  $\text{sr\_all}^* = 0$*

*Sinon, si  $\text{taux1}(i) \times (1 + \text{sr\_comp} + \text{sr\_all}) > 1$ ,  $\text{sr\_comp}^* = \text{sr\_comp}$   
et  $\text{sr\_all}^* = 1/\text{Taux1}(i) - (1 + \text{sr\_comp})$*

*Sinon,  $\text{sr\_comp}^* = \text{sr\_comp}$  et  $\text{sr\_all}^* = \text{sr\_all}$*

La probabilité pour une personne tirée dans l'échantillon de se voir attribuer un questionnaire complet est alors donnée par :

$$P(\text{questionnaire complet}) = 1 + \text{sr\_comp}^* / (1 + \text{sr\_comp}^* + \text{sr\_all}^*)$$

Les autres personnes se sont vues affecter un questionnaire allégé.

### **3.6. L'affectation des personnes échantillonnées à l'échantillon principal ou à la réserve d'échantillon**

L'échantillon et les extensions d'échantillon sont calibrés pour atteindre des cibles sur le nombre de répondants attendus. Il ne s'agit toutefois pas d'une enquête par quotas : tous les individus échantillonnés doivent être interrogés, même si les cibles sont dépassées.

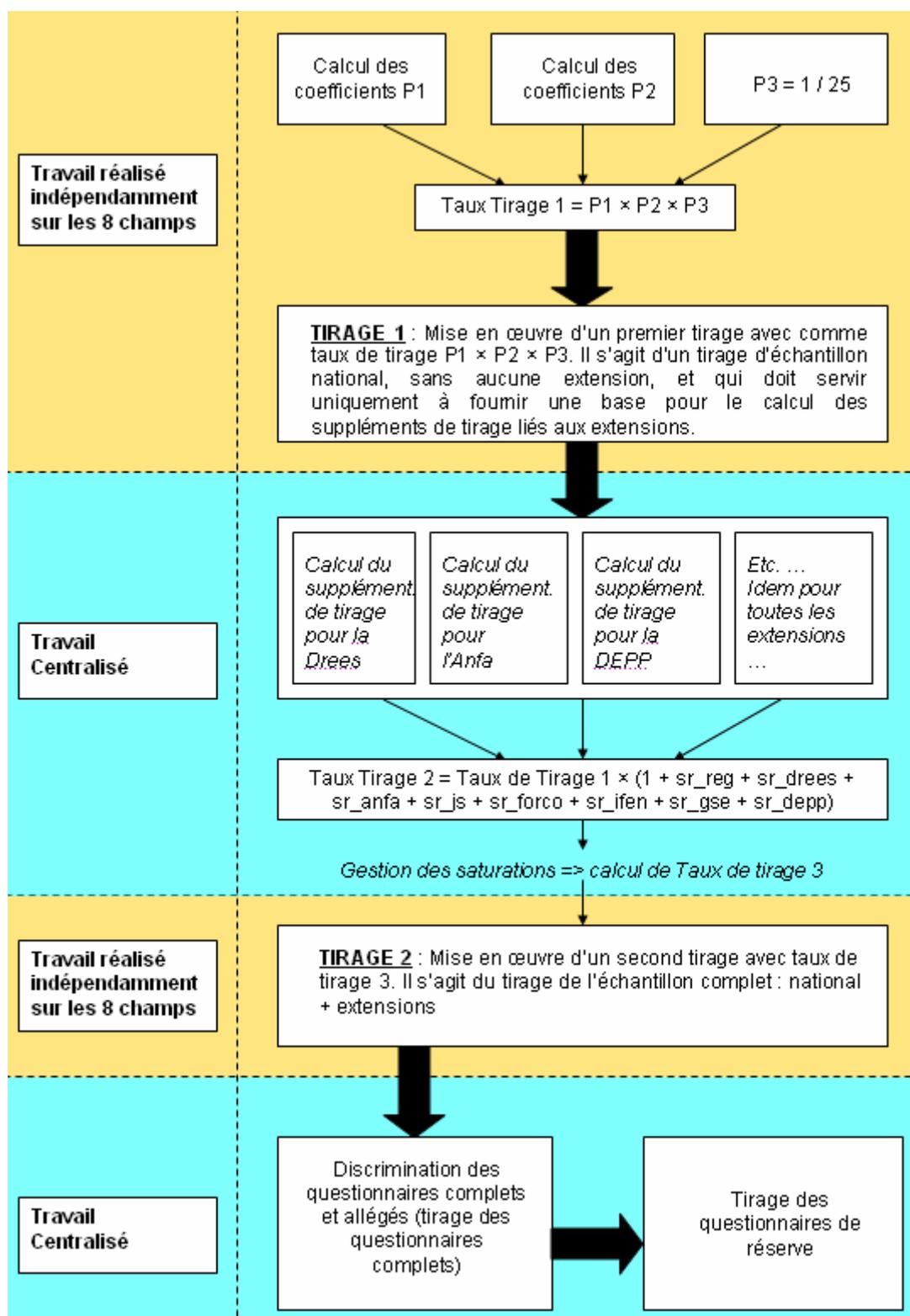
La contrainte budgétaire oblige toutefois à contrôler au mieux le nombre de réponses effectives à l'enquête pour éviter une explosion des coûts dans le cas où le taux de réponse ou le taux de hors champs seraient plus favorables que les niveaux retenus pour calibrer l'échantillon. Une façon de gérer ce risque consiste à répartir les personnes échantillonnées entre un échantillon principal et une ou plusieurs réserves d'échantillon.

Au départ, seul l'échantillon principal est mis en collecte. Il doit être intégralement enquêté. S'il apparaît en cours de collecte que l'échantillon principal ne suffit pas à atteindre les cibles souhaitées sur le nombre de répondants, les échantillons de réserve sont libérés et doivent alors être intégralement enquêtés. Cette manière de procéder permet de conserver le caractère aléatoire de l'échantillon, nécessaire pour calculer des intervalles de confiance.

Dans le cas de l'enquête « Génération 2004 à trois ans », l'échantillon total a été réparti entre un échantillon principal représentant 90% du total et un échantillon de réserve représentant les 10% restant. La répartition s'est faite aléatoirement.

L'échantillon final total est de 403 476 individus, réserve incluse.

Figure 12. Schéma du processus d'échantillonnage



## **4. LA COLLECTE**

---

La collecte de l'enquête s'est déroulée par téléphone d'avril à juin 2007. Un cinquième de la réserve a dû être mobilisé en complément de l'échantillon principal, soit environ 371 500 individus au total. Aucun numéro de téléphone n'avait été récupéré suite aux recherches téléphoniques pour 84 000 jeunes de l'échantillon, soit 22 % et 172 000, pour lesquels au moins un numéro de téléphone était disponible, n'ont pu être joint ou ont refusé de répondre. Au total, 106 000 personnes ont donc pu être jointes et ont accepté de répondre à l'enquête. Parmi elles, environ 50 000, soit 44 %, se sont avérées hors champ. Le nombre total de répondants s'élève à 65 000.

L'enquête s'effectuant par téléphone, la capacité à disposer de coordonnées téléphoniques fiable est essentielle. Dans 73 % des cas, les coordonnées téléphoniques n'étaient pas disponibles dans les fichiers transmis par les établissements de formation. Les autres pouvaient avoir changé de numéro de téléphone, notamment en cas de déménagement au cours des trois années qui séparent la sortie de formation de la période d'enquête. La phase d'enrichissement de l'échantillon par les coordonnées téléphoniques, réalisée en amont de la collecte, est donc cruciale. Elle peut-être source de biais dans les résultats si le profil des personnes pour lesquelles on ne dispose pas de numéro de téléphone valide diffère fortement des autres. Un travail exploratoire sur ce sujet est présenté dans l'annexe 6.

L'enrichissement de l'ensemble de l'échantillon (y compris les réserves) par les coordonnées téléphoniques a permis de retrouver un écho unique, donc un numéro de téléphone unique, pour 36 % des individus, plusieurs échos exploitables dans 31 % des cas. 33 % des individus n'ont pas été retrouvés ou présentaient trop d'échos mais parmi ces derniers, certains disposaient néanmoins d'un numéro de téléphone dans la base de sondage.

Pour augmenter les chances de retrouver les individus échantillonnés, les critères de recherche téléphoniques étaient souples, ne se basant pas strictement sur l'identité et l'adresse de résidence à la date de sortie, de façon à avoir une chance de retrouver directement la personne en cas de déménagement proche de son ancien lieu de résidence ou de retrouver un proche en mesure de donner ses coordonnées. Pour cette raison, certains individus ont été contactés une première fois préalablement à l'enquête, pour s'assurer de leur identité et vérifier en même temps leur appartenance au champ de l'enquête. 91 700 individus ont été concernés par cette phase de pré qualification. Parmi eux, 23 300 ont ensuite été appelés par un plateau d'enquête.

### **4.1. L'enrichissement des coordonnées téléphoniques**

La base de sondage comportait des numéros de téléphones dans 27 % des cas. L'enrichissement de l'échantillon par les coordonnées téléphoniques avait pour triple objectif de compléter les lacunes de la base, d'actualiser le cas échéant le numéro de téléphone des personnes qui avaient déménagé et de trouver, le cas échéant, le numéro de téléphone d'un proche de la famille de l'individu. Il s'agissait donc d'augmenter les chances de joindre les individus présents dans l'échantillon, éventuellement par l'intermédiaire de leur famille proche.

L'enrichissement a été confié après appel d'offre à une société externe disposant des informations présentes dans l'annuaire de France Telecom, seul annuaire disponible pour des recherches téléphoniques en l'absence d'annuaire de mobiles et numéros spéciaux comme les lignes dégroupées<sup>14</sup>. De ce fait, d'ailleurs, un léger biais est possible si les profils des personnes ne disposant que d'un mobile ou d'un numéro spécial diffèrent fortement des autres.

#### ***4.1.1. Le protocole de recherche***

Un protocole de recherche par étapes a été retenu, après un test dont les résultats ont été jugés encourageants.

Dans une première phase, dite « phase A », l'objectif était de retrouver le jeune lui-même ou à défaut ses parents, en recherchant les individus pour lesquels les nom, prénom et adresse aboutissaient à un seul écho, en acceptant un relâchement léger des critères de recherche : légers écarts d'orthographe sur le nom, le prénom ou l'adresse, voire en l'absence de l'adresse ou du prénom.

Si cette première phase n'aboutissait pas à un écho unique, une deuxième phase était mise en œuvre pour l'individu, en relâchant davantage les critères de recherche, dans l'objectif de récupérer au moins des contacts avec des tiers, membres de la famille du jeune, avec en contrepartie, un risque accru d'aboutir à des homonymes :

- soit en relâchant le critère géographique pour rechercher des échos sur les critères « nom et prénom et commune » ou « nom et prénom et département », en autorisant là encore des légers écarts d'orthographe ;
- soit en supprimant tout recours au prénom pour rechercher des échos sur les critères « nom et commune » et « nom et département de l'adresse ».

Ces deux types de recherche sont par la suite respectivement désignés sous les termes de « phase C » et « phase B », même si elles ont été mises en œuvre simultanément. La phase B, sans le recours au prénom est une nouveauté par rapport aux enquêtes « Génération » antérieures. A l'issue de cette seconde étape, un fichier individuel indiquant le nombre d'échos par type de recherche a été récupéré, les numéros récupérés par la phase C étant remontés prioritairement à ceux issus de la phase B, les numéros associés à la commune de résidence à la date de sortie étant également prioritaires par rapport aux autres numéros. Au plus cinq numéros ont été retenus, les individus associés à trop d'échos (plus de dix) étant par ailleurs comptabilisés parmi les échecs à l'issue de cette deuxième étape.

#### ***4.1.2. Bilan de l'enrichissement***

Le Céreq a envoyé en recherche les 403 500 personnes de l'échantillon total, dont 25 000 appartenant à des réserves qui n'ont finalement pas été utilisées.

En se restreignant aux individus appartenant à l'échantillon principal ou aux réserves mobilisées, un écho unique a pu être retrouvé sur les critères « Nom exact, adresse exacte, prénom différent, une seule réponse » dans 26% des cas. Dans 10% des cas, un écho unique a été obtenu en relâchant l'un

---

<sup>14</sup> La proportion de détenteurs exclusifs de portable était alors estimée à 9% par l'Arcep.

des critères. Un numéro de téléphone unique a donc été récupéré à l'issue de la phase A pour 36% des individus échantillonnés.

Des coordonnées téléphoniques ont été récupérées uniquement par la phase B pour 27% des individus et uniquement par la phase C pour 3% des individus. Enfin, pour 2% des individus, des coordonnées téléphoniques ont été récupérées dans les deux phases B et C.

Pour un tiers des individus, les recherches téléphoniques n'ont donc donné aucun résultat : aucun écho ou échos trop nombreux. Parmi ces derniers, 31% disposaient néanmoins d'un numéro de téléphone dans la base de sondage. Au total, il restait donc 22% individus échantillonnés sans coordonnée téléphonique à l'issue de la phase d'enrichissement, au lieu de 73% avant celle-ci.

**Figure 13. Bilan de la phase d'enrichissement en numéros de téléphone (données arrondies)**

	Numéro de téléphone disponible dans la base de sondage	Pas de numéro de téléphone disponible dans la base de sondage	
Numéro récupéré en phase A (écho unique)	13 400	121 100	134 500
Un seul numéro récupéré en phase B ou C	14 400	29 700	44 100
Plusieurs numéros récupérés en phase B ou C	23 100	48 100	71 200
Aucun numéro récupéré lors de la phase d'enrichissement	37 900	83 830	121 730
	88 800	282 730	371 530
Champ : ensemble de l'échantillon principal et des réserves utilisées lors de l'enquête.			

La pertinence de la phase d'enrichissement ne se juge toutefois pas à partir de la réduction de cet écart mais dans la capacité à trouver le bon numéro de téléphone pour joindre effectivement ensuite la personne. La figure 14 permet d'approcher le rendement de la phase d'enrichissement sous deux angles : le nombre moyen de numéros récupérés qui ont été utilisés en enquête et le taux d'enquêtes effectivement réalisées, en se restreignant aux 371 500 individus échantillonnés transmis en plateau d'enquête (échantillon principal et réserves utilisées).

Pour les 37 900 individus qui n'ont pas été retrouvés dans la phase d'enrichissement mais pour lesquels un numéro de téléphone était disponible dans les fichiers transmis par les établissements, le taux d'enquêtes réalisées atteint 20 %, c'est-à-dire que 20 % des personnes ont pu être contactées, ont accepté de répondre et entraient dans le champ de l'enquête. En appliquant ce taux de succès aux 23 % des individus échantillonnés de la base de sondage pour lesquels la base de sondage contenait un numéro de téléphone, le potentiel de réponse en l'absence de phase d'enrichissement pourrait être estimé à environ 20 000 répondants, soit trois fois moins que le nombre de répondants finalement obtenus dans l'enquête.

**Figure 14. Taux de succès en fonction du nombre de numéros récupérés** (données arrondies)

Nombre de numéros récupérés lors de la phase d'enrichissement	Nombre d'individus échantillonnés concernés	Nombre moyen de numéros récupérés utilisés en plateau d'enquête	Taux d'enquêtes réalisées (répondants entrant dans le champ)
0	121 700	0	6,3
<i>(dont ceux ayant un numéro de téléphone dans la base de sondage)</i>	<i>(37 900)</i>	<i>(0)</i>	<i>(20,2)</i>
1	178 600	0,7	27,6
2	27 700	1,1	12,2
3	18 500	1,4	11,8
4	13 700	1,8	11,4
5	11 300	2	11,1
Total	371 500		17,6

Champ : ensemble de l'échantillon principal et des réserves utilisées lors de l'enquête.

Note de lecture : pour 27 697 individus, les recherches téléphoniques ont permis l'enrichissement de deux numéros de téléphone. En moyenne et par individu, le plateau d'enquête en a utilisé 1,1. Au final, 12,2% de ces individus ont répondu à l'enquête.

Lorsque la phase d'enrichissement donne un écho unique, donc notamment la phase A de la recherche de coordonnées téléphoniques, le taux d'enquêtes réalisées augmente d'un tiers pour approcher 28 %, soulignant l'intérêt de confirmer et d'actualiser les numéros de téléphone contenus dans la base de sondage. Lorsque la phase d'enrichissement en coordonnées téléphoniques fournit plusieurs numéros de téléphone, le taux d'enquêtes réalisées chute et diminue avec le nombre de numéros de téléphone récupérés, passant de 12,2 % quand deux numéros ont été récupérés à 11,1 % quand cinq numéros ont été récupérés. En réalité, d'ailleurs, le nombre de numéros de téléphones récupérés dans la phase d'enrichissement qui sont réellement utilisés varie beaucoup moins puisqu'il s'établit à 1,1 quand deux numéros ont été récupérés et atteint deux quand cinq numéros ont été récupérés. Cette faible variation conforte la pertinence des ordres de priorité fixés lors du classement des numéros de téléphone récupérés. Elle ne signifie pas pour autant que le nombre de numéros de téléphone récupérés lors de la phase d'enrichissement pourrait être borné avec un seuil inférieur à cinq. Il faudrait, pour cela, poursuivre les analyses en étudiant la dispersion du nombre de numéros mobilisés et les profils des personnes échantillonnées et des personnes répondantes, selon le nombre de numéros de téléphone mobilisés.

L'annexe 6 propose par ailleurs un bilan qualitatif de l'introduction de la phase B dans la recherche de coordonnées téléphoniques. Il montre que les personnes contactées grâce à un numéro récupéré dans cette phase, sur la base du nom et de la commune ou du département de résidence à la date de fin d'études, sans référence au prénom, ni à l'adresse précise, vivent moins souvent seuls et plus souvent chez leurs parents que les autres ; ils sont moins diplômés et plus souvent au chômage à la date d'enquête.

### **Encadré 5 - Des améliorations possibles ?**

La méthode retenue pour l'enrichissement de l'échantillon en coordonnées téléphoniques rencontre des limites. La principale résulte du champ de la recherche. Les téléphones mobiles et les fournisseurs d'accès Internet qui sont maintenant détenteurs des lignes téléphoniques ne sont pas pris en compte puisqu'ils n'étaient pas dans l'annuaire de France Télécom. Aucune solution n'a été trouvée pour résoudre cette difficulté, en l'absence d'entreprises disponibles sur le marché pour tenir compte de ce segment.

Par ailleurs, l'utilisation du nom de la commune plutôt que du code postal, conduit parfois à des recherches géographiques moins précises. Mais, a contrario, dans certains cas où le nom du bureau distributeur est saisi dans la base sondage à la place du nom de la commune, alors que le bureau distributeur regroupe plusieurs communes, la recherche se centre alors sur une commune qui n'est pas forcément la bonne.

Enfin, les numéros envoyés en plateau d'enquête sont rarement tous utilisés (au maximum 6 dont le numéro de téléphone de la base de sondage). La qualification se fait principalement sur le premier numéro. Plus le nombre d'échos est faible, plus on retrouve les individus.

## **4.2. La phase de pré qualification**

La phase d'enrichissement en coordonnées téléphoniques conduit à disposer pour certains individus de plusieurs numéros de téléphone susceptibles d'être le leur ou celui d'un de leurs proches. La phase de pré qualification a pour but d'aboutir à un numéro de téléphone unique à enquêter, grâce à une validation préalable de la personne, voire une vérification de l'appartenance au champ de l'enquête pour les individus relevant de formations où les poursuites d'études sont nombreuses.

A cet effet, à l'issue de la phase d'enrichissement en coordonnées téléphoniques, tous les individus présents dans l'échantillon ont été classés selon le degré de fiabilité présumé des coordonnées téléphoniques récupérées et le risque présumé d'être hors champ. Puis, selon les cas, les individus ont été orientés vers une phase de pré qualification destinée à valider l'identité de la personne et son appartenance au champ (25 % des cas) ; ils ont été directement envoyés en enquête (53 % des cas) ou définitivement écartés de l'enquête, faute de coordonnées téléphoniques fiables (22 %).

La pré-qualification a porté à la fois sur l'échantillon principal et sur les échantillons de réserve libérés en cours de collecte. Parmi les 91 700 individus traités en pré-qualification, 26 % ont finalement été orientés vers l'enquête ; 25 % ont été jugés hors cible (l'individu contacté n'était pas celui recherché) et 12 % étaient hors champ (l'individu contacté était bien celui recherché mais ne faisait pas partie du champ de l'enquête). Dans 20 % des cas, aucun contact n'a pu être établi (faux numéro, répondeur, non réponse) et dans 18 % des cas, le contact a été infructueux (refus de réponse de l'individu ou du tiers répondant, individu à l'étranger, décédé, ...).

### ***4.2.1. La détermination des individus concernés par la phase de pré qualification***

Après la phase d'enrichissement en coordonnées téléphoniques et avant l'envoi en enquête, les individus de l'échantillon principal et des réserves utilisées ont été classés en fonction de la fiabilité

des coordonnées téléphoniques récupérées et du risque que l'individu soit hors champ, pour être répartis suivant les cinq cas décrits dans le tableau ci-dessous.

Au total, 91 700 individus, soit un quart de l'échantillon et des réserves mobilisées ont été envoyés en pré-qualification :

- 28 200 parce qu'ils appartenaient à des strates de formation présentant habituellement les taux de hors champ les plus élevés (plus de 80%), soit en pratique les sortants d'établissements privés universitaires de niveau III, de 4<sup>e</sup> technologique ou avec dispositif aménagé, de 3<sup>e</sup> (générale, technologique ou d'insertion) de seconde, de première L et S, de terminale L et S, de baccalauréat technologique ou S en lycée agricole, de CAP, de classe d'initiation pré-professionnelle en alternance (« cas 0 ») ;

- 63 500 parmi les autres strates, parce que plusieurs numéros de téléphone ont été récupérés à l'issue des phases d'enrichissement en coordonnées téléphoniques B et C (« cas 3 »).

Dès lors qu'ils avaient des coordonnées téléphoniques (« cas 1 » et « cas 2 »), les individus ont été directement envoyés en enquête, tandis que ceux pour lesquels aucune coordonnée téléphonique n'était disponible (« cas 4 ») étaient définitivement écartés de la collecte de l'enquête.

**Figure 15. L'orientation vers la phase de pré-qualification ou non selon les cas**

<b>Cas 0</b>	Individus avec des coordonnées téléphoniques appartenant aux strates de formation présentant habituellement les taux de hors champ les plus élevés (plus de 80%)	<b>8%</b> des individus de l'échantillon principal et des réserves libérées étaient dans cette situation.  ⇒ <i>Passage en pré qualification</i>
<b>Cas 1</b>	Individus appartenant aux strates de formation présentant habituellement les taux de hors champ inférieurs à 80%, pour lesquels la phase A de la recherche des coordonnées téléphoniques a abouti à un numéro de téléphone unique ou qui n'ont pas été retrouvés lors des recherches téléphoniques mais qui possédaient un numéro de téléphone fourni par l'établissement de formation dans la base de sondage.	<b>42%</b> des individus  ⇒ <i>Passage direct A l'enquête</i>
<b>Cas 2</b>	Individus appartenant aux strates de formation présentant habituellement les taux de hors champ inférieurs à 80%, pour lesquels les phases B et C de la recherche des coordonnées téléphoniques ont abouti à un numéro de téléphone unique.	<b>11%</b> des individus  ⇒ <i>Passage direct A l'enquête</i>

<b>Cas 3</b>	Individus appartenant aux strates de formation présentant habituellement les taux de hors champ inférieurs à 80% pour lesquels les recherches téléphoniques ont conduit à des numéros multiples mais limités à cinq numéros maximum lors des recherches élargies (phases B et C).	<b>17%</b> des individus  ⇒ <i>Passage en pré qualification</i>
<b>Cas 4</b>	Individus qui ne possèdent pas de numéro de téléphone dans la base de sondage et pour lesquels les recherches téléphoniques ont échoué ou ont conduit à des échos multiples trop nombreux (plus de 5 numéros).	<b>22%</b> des individus  ⇒ <i>Non enquêtés</i>

#### ***4.2.2. Le déroulement de la phase de pré qualification***

La pré-qualification s'est déroulée entre le 12 mars et le 30 juin 2007, par téléphone, à l'aide d'un questionnaire assisté par informatique, selon les mêmes créneaux horaires que l'enquête : en journée (11h45-17h45) et en soirée (18h-20h45), du lundi au vendredi et le samedi, de 10h à 18h. Cinq à vingt-deux télé-enquêteurs constituaient l'équipe de jour en semaine. Cinq à vingt-cinq télé-enquêteurs constituaient l'équipe du soir en semaine et du samedi.

Six numéros de téléphone (y compris éventuellement celui issu de la base de sondage) étaient au maximum disponibles par individu. Ils étaient classés selon l'ordre de priorité suivant : numéro issu de la phase A de recherche, numéro issu de la base de sondage, numéros issus de la phase C de recherche, puis numéros issus de la phase B de recherche.

Pour un individu donné, le numéro le plus prioritaire était d'abord composé. Dans le cas où ce numéro s'avérait infructueux (réponse mais personne hors cible ; faux numéro ; quinze appels successifs à des heures et des jours différents sans réponse), le numéro suivant était composé et ainsi de suite jusqu'à la qualification de l'individu. Au terme du nombre maximum programmé de tentatives d'appels sur l'ensemble des numéros, la fiche était classée dans la rubrique « abandon après maximum d'appels ».

Un automate d'appels a été utilisé au début de la phase de pré-qualification, au regard de l'important volume d'enregistrements à traiter. Il permettait un traitement optimisé des différents rangs des téléphones des individus. L'automate composait de façon automatique les numéros de téléphone et transférait aux télé-enquêteurs seulement les appels aboutis accompagnés de leur fiche informatique. Lorsque les volumes de téléphones ont été réduits, une composition manuelle des numéros de téléphone a remplacé l'automate d'appels.

Une fois l'appel décroché, le télé-enquêteur déroulait un questionnaire « Cati »<sup>15</sup> pour valider ou non l'identité de la personne. Lorsque la personne répondant à l'appel était l'individu cherché, le télé-

---

<sup>15</sup> « Computer assisted telephone interview », ou enquête téléphonique assistée par ordinateur. Deux « Cati » ont été développés pour la phase de pré-qualification, l'un pour les individus concernés par le questionnaire complet, l'autre pour les individus concernés par le questionnaire allégé. Les développements de ces questionnaires sous informatique ont été préalablement testés entre le 12 et le 17 février 2007, à partir de 2 021 fiches tests issus de la base de sondage de l'enquête ,

enquêteur validait aussi son appartenance au champ de l'enquête en lui posant la partie « filtre » qui ouvre le questionnaire de l'enquête, dans ses deux versions (allégé ou complet). En revanche, dans tous les cas, la passation éventuelle du questionnaire était réalisée plus tard.

De façon plus précise, après vérification de l'identité de l'individu :

- Si la personne en ligne était l'individu recherché, le questionnaire filtre était posé, aboutissant à la qualification de la personne, à sa mise hors champ ou à un refus de réponse.
- Si la personne en ligne était un homonyme parfait ou un tiers ne connaissant pas l'individu, le numéro composé était classé hors cible et le numéro suivant était testé.
- Si la personne en ligne était un tiers acceptant de transmettre un nouveau numéro, celui-ci était ajouté à la base. Ce numéro était ensuite testé en priorité dans un appel suivant de pré-qualification pour les individus relevant du cas 0 pour vérifier l'appartenance au champ de la personne. Il était directement envoyé en terrain d'enquête pour les individus relevant du cas 3.
- Si la personne en ligne était un tiers qui connaissait l'individu mais pas ses coordonnées, le numéro composé était classé hors cible et le numéro suivant était testé.

#### ***4.2.3. Bilan de la phase de pré qualification***

Parmi les 91 737 individus envoyés en pré-qualification en début et au cours du terrain, 73 520 ont donné lieu à au moins un contact :

- 4 712 ont été envoyés en enquête après qualification (validation du questionnaire filtre),
- 18 776 ont été envoyés directement en enquête après validation du numéro par un tiers (donc sans validation du questionnaire filtre),
- 12 045 ont été classés hors champs (l'individu contacté était bien celui recherché mais ne faisait pas partie du champ de l'enquête)
- 1 310 ont refusé de répondre,
- 43 avaient déjà répondu à l'enquête,
- 326 ont été déclarés décédés par un tiers,
- 582 ont été déclarés résidant à l'étranger par un tiers,
- 5 712 ont été déclarés injoignables permanents,
- 12 274 ont été classés sans suite à cause d'un refus de tiers,
- 29 528 ont été classés hors cible (l'individu contacté n'était pas celui recherché).

6 429 n'ont pas pu être joints (faux numéro, ne répond pas, occupé, répondeur,...).

Concernant cette fois les numéros de téléphone, ce sont environ 145 400 numéros de téléphone qui ont été traités lors de la phase de pré-qualification. Parmi eux, 74 800 ont été classés hors cible, soit 60%, et 14 100, soit 10%, ont donné lieu à un abandon après un maximum d'appels.

Pour les personnes concernées, la passation du questionnaire filtre de qualification a duré en moyenne 3 mn 30.

---

dont 89 (4%) ont été qualifiées. Ces tests en situation réelle ont permis de valider l'exactitude des programmes, de vérifier la bonne compréhension de l'ensemble des questions et d'obtenir une estimation du minutage.

Dans 9,6% des cas, la passation du questionnaire a été réalisée sur un téléphone portable et dans 4,3% des cas sur un numéro spécial.

Les taux de contact obtenus en journée ont été légèrement inférieurs à ceux enregistrés en soirée ou le samedi.

La figure 16 présente les résultats détaillés de la phase de pré qualification au niveau des individus, en distinguant les quatre terrains de l'opération construits pour distinguer les personnes selon qu'elles étaient ou non concernées par le questionnaire complet et selon le cas à l'origine de l'orientation vers la pré qualification (« cas 0 » ou « cas 3 »).

A l'issue de la pré-qualification, il s'avère beaucoup plus difficile d'établir le contact pour les individus relevant du « cas 0 » que pour ceux relevant du « cas 3 ». Dans 30% des cas, il n'y a aucun contact pour les premiers (« abandon après le nombre maximum d'appel » ou « individus non classés »), soit deux fois plus que pour les seconds. Les individus relevant du cas 0 sont en effet souvent issus de collèges et lycées, pour lesquels aucun numéro de téléphone n'était disponible dans la base de sondage. Il en est ainsi pour 98% d'entre eux et pour 72%, un seul numéro de téléphone était disponible après la phase d'enrichissement. Au contraire, pour les individus relevant du cas 3, les numéros de téléphone sont multiples et comportent dans 36% des cas, un numéro issu de la base de sondage.

Toujours dans le « cas 0 », il y a beaucoup plus souvent de contact direct avec l'enquêté puisque le protocole retenu interdisait d'envoyer directement en enquête les individus concernés sans leur avoir préalablement posé le questionnaire filtre pour vérifier leur appartenance au champ de l'enquête.

**Figure 16. Bilan de la phase de pré-qualification, par individu**

Résultats par individu	Allégé				Complet				Total	
	Cas 0 : CFA1	%	Cas 3 : CFA2	%	Cas 0 : CFC1	%	Cas 3 : CFC2, CFC3	%	Cas 0 + 3	%
Qualification	477	6,9%	1 326	5,2%	1 128	5,3%	1 781	4,7%	4 712	5,1%
Envoi en enquête	-	0,0%	8 210	32,1%	-	0,0%	10 566	27,8%	18 776	20,5%
Hors champ	2 580	37,3%	1 490	5,8%	6 661	31,2%	1 314	3,5%	12 045	13,1%
Refus de l'individu	89	1,3%	378	1,5%	284	1,3%	559	1,5%	1 310	1,4%
A déjà répondu	3	0,0%	8	0,0%	14	0,1%	18	0,0%	43	0,0%
Décédé	22	0,3%	84	0,3%	70	0,3%	150	0,4%	326	0,4%
Réside à l'étranger	52	0,8%	213	0,8%	96	0,5%	221	0,6%	582	0,6%
Injoignable permanent	1 184	17,1%	216	0,8%	3 926	18,4%	386	1,0%	5 712	6,2%
Refus d'un tiers	680	9,8%	3 570	14,0%	1 853	8,7%	6 171	16,3%	12 274	13,4%
Hors cible	1 074	15,5%	9 108	35,6%	4 138	19,4%	15 208	40,1%	29 528	32,2%
Pas d'interlocuteur, ne connaît pas la cible	753	10,9%	948	3,7%	3 146	14,8%	1 582	4,2%	6 429	7,0%
<b>Total</b>	<b>6 914</b>	<b>100,0%</b>	<b>25 551</b>	<b>100,0%</b>	<b>21 316</b>	<b>100,0%</b>	<b>37 956</b>	<b>100,0%</b>	<b>91 737</b>	<b>100,0%</b>

Champ : Ensemble des individus concernés par la pré qualification.

Note : Le terrain d'opération « CFC1 » concerne les individus relevant du « cas 0 » et destinés au questionnaire complet. Le terrain d'opération « CFC2 » concerne les individus relevant du « cas 3 » et destinés au questionnaire complet. Le terrain d'opération « CFA1 » concerne les individus relevant du « cas 0 » et destinés au questionnaire allégé. Le terrain d'opération « CFA2 » concerne les individus relevant du « cas 3 » et destinés au questionnaire allégé.

### **4.3. L'amélioration des phases d'enrichissement et de pré qualification par rapport aux enquêtes précédentes**

#### *4.3.1. Nature des évolutions*

Pour les enquêtes « Génération 1998 » et « Génération 2001 », les recherches téléphoniques se sont effectuées pendant la phase d'enrichissement à partir des nom et prénom de l'individu, par élargissement géographique progressif à partir de l'adresse initiale : département, puis région, puis Ile-de-France et reste de la France. Pendant la phase de pré qualification, jusqu'à trente numéros de téléphone ont pu être retenus et testés avant de trouver le bon individu parmi les différents homonymes, le nombre maximum de tentatives d'appels étant limité à cinq par numéro.

Pour l'enquête « Génération 2004 », des évolutions importantes ont été apportées. Un nouveau type de recherche de coordonnées téléphoniques a été mis en œuvre dans la phase d'enrichissement (la phase B présentée plus haut), sans utiliser le prénom, donc uniquement à partir du nom de l'individu, d'abord dans la commune de l'adresse puis dans le département, dans l'idée d'atteindre des membres de la famille du jeune susceptibles de fournir son numéro de téléphone et d'améliorer ainsi le taux de réponse, mais aussi avec le risque d'aboutir plus souvent à un homonyme.

Cette évolution a conduit à faire évoluer le contenu même de la phase de pré qualification. Jusque là construite autour des individus relevant des « cas 0 », c'est-à-dire des strates de formation pour lesquelles la base de sondage contient beaucoup de personnes hors champ, elle avait pour rôle de filtrer les personnes à envoyer en enquête. Pour l'enquête « Génération 2004 », la phase de pré qualification est aussi devenue une phase de recherche d'individus, avec possibilité de récupération de coordonnées téléphoniques.

Le protocole de la phase de pré-qualification a également été revu, avec une réduction drastique du nombre de numéros de téléphone récupérés dans la phase d'enrichissement, au profit d'une sélection sur la base d'une hiérarchisation des numéros retrouvés et d'une augmentation du nombre d'appels possibles en cas d'appel non décroché.

**Figure 17. Présentation synthétique des changements introduits pour « Génération 2004 »**

	<b>Génération 1998 &amp; Génération 2001</b>	<b>Test de Génération 2004</b>	<b>Génération 2004</b>
Public concerné	<ul style="list-style-type: none"> <li>- Individus issus de formations à fort taux de hors champ</li> <li>- Individus issus de recherches élargies avec plusieurs N° de tél</li> <li>- Individus non retrouvés lors de recherches élargies mais possédant un N° issu de la base de sondage</li> </ul>	Individus non retrouvés sur adresse et issus de recherches élargies	<ul style="list-style-type: none"> <li>- Taux de hors champs élevé</li> <li>- Individus issus de la recherche élargie à échos multiples</li> </ul>
Méthode de Qualification	Passage du questionnaire filtre obligatoire	Passage du questionnaire filtre obligatoire	<p>2 possibilités :</p> <ul style="list-style-type: none"> <li>- Basculement terrain si récupération des coordonnées de l'individu cherché (sauf cas 0)</li> <li>- Questionnaire filtre si l'enquêteur est directement en contact avec l'individu ou si cas 0</li> </ul>
Méthode de recherche téléphonique	<ul style="list-style-type: none"> <li>- Phase A (individu retrouvé sur adresse avec un code de fiabilité jugé suffisant)</li> <li>- Recherches élargies sur nom+prénom France : département, région, Ile de France, France</li> </ul>	<ul style="list-style-type: none"> <li>- Phase A</li> <li>- Elargissement Nom Prénom Commune (NPC)</li> <li>- Elargissement Nom Prénom Département (NPD)</li> <li>- Elargissement Nom Commune (NC)</li> <li>- Elargissement Nom Département (ND)</li> <li>- Elargissement département de résidence des parents si différent.</li> </ul>	<ul style="list-style-type: none"> <li>- Phase A</li> <li>- Elargissement NPC</li> <li>- Elargissement NPD</li> <li>- Elargissement NC</li> <li>- Elargissement ND</li> </ul>
Nombre de téléphones maxi récupérés	30 numéros	6 numéros	5 numéros + numéro de la base de sondage
Automate d'appels	Oui	Non	Oui
Taux d'insistance	<p>10 tentatives par numéro quand plusieurs numéros (revu à 5 en cours de terrain)</p> <p>30 tentatives pour les N° uniques</p>	Gestion individualisée des fiches	15 tentatives

### **4.3.2. Le bilan de deux tests préalables**

Deux tests successifs ont été réalisés afin de valider la faisabilité et la pertinence de ces évolutions.

Un premier test papier a été réalisé en interne auprès d'un échantillon de 782 individus non retrouvés avec une recherche de coordonnées téléphoniques s'appuyant sur le nom, le prénom et l'adresse. Ce premier test a permis de valider un plan de dialogue adapté aux nouveaux objectifs de la pré-qualification et de montrer qu'une récupération des bonnes coordonnées téléphoniques était possible.

Un second test a été effectué en conditions réelles, avec un questionnement assisté par ordinateur, lors du test de questionnaire réalisé sur un échantillon de 8 000 jeunes. A cette occasion, des recherches élargies (phases B et C) ont été effectuées pour 4 900 individus, permettant de tester la nouvelle opération de pré-qualification auprès de 2 000 individus. Lors de ce test, les numéros de téléphone étaient classés selon l'ordre de priorité suivant : nom et prénom de l'individu sur la commune de résidence (échos multiples), nom et prénom de l'individu sur le département de résidence, nom de l'individu sur la commune de résidence, nom de l'individu sur le département de résidence, nom de l'individu sur le département de résidence des parents (si différent). Le test a notamment permis de qualifier une centaine d'individus (4,6 %), d'identifier plus de 300 individus hors cible (17 %) et environ 70 hors champ (3,4 %). 1 400 individus n'ont pas été classés (69,9 %).

*In fine*, on estime que la mise en œuvre de ces nouvelles recherches téléphoniques a permis d'augmenter de 3 points le taux de réponse à l'enquête, soit une hausse de plus de 10 %.

## **4.4. La collecte**

### **4.4.1. L'organisation générale de la collecte**

La collecte de l'enquête s'est déroulée du 2 avril au 12 juillet 2007, par téléphone, à l'aide d'un questionnement assisté par ordinateur, sur trois créneaux horaires : en journée (11h45-17h45) et en soirée (18h-20h45), du lundi au vendredi et le samedi, de 10h à 18h. Deux équipes ont été constituées : l'équipe jour (semaine en journée), avec un effectif moyen de quatre-vingt-cinq télé-enquêteurs, et l'équipe du soir et du samedi, avec un effectif moyen variant de cent quatre télé-enquêteurs en soirée à cent dix-neuf télé-enquêteurs le samedi.

L'enquête est assisté par ordinateur, donc l'enquêteur interroge les individus par téléphone tout en suivant sur un écran d'ordinateur un script préétabli qui affiche les questions qu'il doit poser et les éventuelles modalités de réponse, avec la consigne de lire strictement les questions et de lister les modalités, sauf mention contraire<sup>16</sup>. Les réponses sont saisies directement sur informatique. Le Cati

---

<sup>16</sup> Deux Catis ont été programmés pour l'enquête, l'un pour le questionnaire complet et l'autre pour le questionnaire allégé. Les programmes ont été conçus par la société prestataire retenue pour réaliser les enquêtes. Ils ont été testés et validés en bureau par le Céreq puis testés par le prestataire auprès d'un échantillon de jeunes entrant dans le champ de l'enquête. Ce test en situation réelle avait pour objectifs de confirmer l'exactitude des programmes, de vérifier la bonne compréhension de l'ensemble des questions et d'obtenir une estimation du minutage. Il s'est déroulé sur deux périodes (du 19 au 21 février puis du 28 février au 2 mars 2007). Au total, 324 enquêtes ont été réalisées sur le questionnaire allégé et 294 sur le questionnaire

est capable de gérer des filtres et d'orienter l'individu vers des questions différentes en fonction des réponses précédemment données.

Afin de spécialiser les télé-enquêteurs sur des cibles particulières (selon le diplôme, le type de questionnaire, le résultat des recherches téléphoniques), l'échantillon a été réparti entre trois plateaux d'enquête, subdivisés au total en onze terrains d'enquête :

- le plateau n°1, destiné à l'ensemble des individus *a priori* bien identifié (numéro de téléphone unique) et de strates à taux de hors champ inférieur à 80% (« cas 1 », dans la figure 15) ;
- le plateau n°2, pour les individus pour lesquels l'identification des coordonnées téléphoniques est plus floue (ensemble des individus relevant du « cas 2 » et ceux du « cas 3 » n'ayant pas répondu au questionnaire filtre lors de la pré-qualification) ;
- le plateau n°3, pour tous les individus relevant de strates à taux de hors champs très élevé (« cas 0 ») et ceux ayant répondu au questionnaire filtre lors de la pré-qualification.

La figure 18 présente plus en détail les différents terrains et leur affectation entre ces trois plateaux.

**Figure 18. Présentation des terrains d'enquête**

Code du terrain d'enquête	N° de plateau	Type de questionnaire posé	Cas concernés (selon la typologie de la figure 15)	Niveau de diplôme	Provenance
CA11	1	Allégé	Cas 1	Sans qualification et CAP/BEP non diplômés	Echantillon
CA12	1	Allégé	Cas 1	CAP/BEP diplômés et niveau Bac	Echantillon
CA13	1	Allégé	Cas 1	Bac+2 et plus	Echantillon
CA21	2	Allégé	Cas 2 + Cas 3 n'ayant pas répondu au questionnaire filtre	Tous niveaux	Echantillon +pré-qualification
CA31	3	Allégé	Cas 0 + Cas 3 ayant répondu au questionnaire filtre	Tous niveaux	Pré-qualification
CL11	1	Complet	Cas 1	Sans qualification et CAP/BEP non diplômés	Echantillon
CL12	1	Complet	Cas 1	CAP/BEP diplômés et niveau Bac	Echantillon
CL13	1	Complet	Cas 1	Bac+2 et plus	Echantillon
CL14	1	Complet	Cas 1	Apprentis et sortants santé-social	Echantillon
CL21	2	Complet	Cas 2 + Cas 3 n'ayant pas répondu au questionnaire filtre	Tous niveaux	Echantillon +pré-qualification
CL31	3	Complet	Cas 0 + Cas 3 ayant répondu au questionnaire filtre	Tous niveaux	Pré-qualification

complet, avec une durée moyenne de 20,2 mn pour le premier et 34,9 mn pour le second. Il a conduit à des adaptations mineures des questionnaires.

Pour la majeure partie de l'échantillon, qui n'était pas passée par la phase de pré-qualification, le questionnaire commençait par la partie filtre destinée à confirmer son identité et son appartenance au champ. Après confirmation, le reste du questionnaire était administré immédiatement. Il en était de même pour les personnes dont l'identité avait été validée par un tiers lors de la phase de pré-qualification et qui n'avaient donc pas répondu au questionnaire filtre. Pour les autres, le questionnaire commençait d'emblée après cette partie.

## **Encadré 6 – La sensibilisation des enquêtés**

### ***Une lettre avis***

Les individus à enquêter dont les adresses paraissaient suffisamment fiables ont reçu une lettre-avis dès lors qu'ils n'appartenaient pas à une strate de formation à fort taux de hors champ. Ont été considérées comme fiables les adresses validées par la récupération d'un numéro en phase A d'enrichissement en coordonnées téléphonique et celles des individus pour lesquels seul le numéro de téléphone issu de la base de sondage était disponible (« cas 1 » dans la figure 15). Au total, 159 062 personnes, soit 43% de l'échantillon et des réserves utilisées, se sont vues adresser une lettre avis.

Cette lettre avis a plusieurs fonctions. Elle a d'abord un rôle informatif. Elle cherche ensuite à sensibiliser le futur enquêté pour faciliter son adhésion lors du contact téléphonique. Elle l'incite à préparer le calendrier d'activité des trois années écoulées, élément clé du questionnaire nécessitant un effort de mémoire. Elle propose également un numéro vert et un site internet. Un modèle de lettre avis figure dans l'annexe 4.

### ***Un site internet et un blog « Génération 2004 »***

Un site internet dédié aux personnes échantillonnées a été mis en place, pour informer sur l'enquête en cours mais aussi dans la perspective du suivi de l'échantillon, celui-ci devant être réinterrogé deux ans et quatre ans plus tard. Accessible à partir de la page d'accueil du site du Céreq, il permettait aux jeunes de consulter les résultats des enquêtes précédentes, d'envoyer un courriel au Céreq et de s'exprimer sur un blog au sujet de l'enquête. Quelques copies d'écrans figurent dans l'annexe 4.

### ***Un numéro vert***

Un numéro vert a été mis en place pour permettre aux jeunes de donner de nouvelles coordonnées, de réaliser l'enquête ou de faire part de leur refus d'y répondre. Le numéro était présent sur la lettre-avis. Il était aussi communiqué lors des contacts téléphoniques à des tiers qui refusaient de transmettre les coordonnées de l'individu ou aux proches d'individus difficiles à joindre.

Ce numéro vert a été ouvert du lundi au samedi, entre le 27 mars et le 13 juillet 2007. En dehors des horaires d'ouverture, les jeunes pouvaient laisser un message sur un répondeur. Il a reçu 9 100 appels dont 200 messages.

#### ***4.4.2. Le suivi de la collecte***

Après une formation initiale des responsables d'enquête et des télé-enquêteurs, trois intervenants du Céreq ont assuré une permanence quotidienne pendant la phase terrain pour apporter un appui technique, notamment pour répondre aux questions et former les nouveaux télé-enquêteurs. Il s'agissait aussi pour eux de contribuer à la qualité de l'enquête en assurant des écoutes régulières des télé-enquêteurs, en contrôlant le respect des règles du questionnement et l'homogénéité de traitement des cas particuliers.

La transmission de statistiques quotidiennes et hebdomadaires a permis le suivi de la collecte (suivi de gestion de l'échantillon, des contacts, des enquêtes réalisées mais aussi des réponses à certaines questions de l'enquête). Des réunions téléphoniques hebdomadaires et des réunions mensuelles sur le bilan d'avancement, tant pour la phase de pré-qualification que pour l'enquête, ont également été organisées.

Le pilotage au plus près des réalisations de questionnaires a conduit à libérer certaines réserves pour atteindre les objectifs fixés, notamment pour les extensions. Un total de 5 200 individus de l'échantillon de réserve ont finalement été envoyés en pré qualification ou en enquête (14% de la réserve). Ils appartenaient aux champs suivants : CFA et extensions (ANFA, Sport, FORCO, PACA).

#### ***4.4.3. Le taux de réponse final***

Comme nous l'avons vu, l'échantillon total, ensemble des réserves comprises, comprenait 403 476 individus. L'échantillon opérationnel, composé de l'échantillon principal et des réserves libérées, s'élève à 371 483 individus. Parmi eux, 83 730 individus relevaient du « cas 4 », pour lesquels on ne disposait pas de numéro dans la base de sondage et pour lesquels les recherches téléphoniques n'ont pas été satisfaisantes (aucun numéro retrouvé ou plus de cinq échos).

Ce sont donc 287 753 personnes, soit 77,5% de l'échantillon opérationnel, qui ont été mises en collecte. Parmi elles, compte tenu des impossibilités à joindre la cible et des refus, seules 115 813<sup>17</sup> ont finalement répondu à la première partie du questionnaire, c'est-à-dire au questionnaire filtre visant à déterminer si l'individu fait bien partie du champ de l'enquête. Parmi eux, 50 424 individus étaient hors champ et 65 389 ont répondu à l'enquête.

Si l'on rapporte ces 115 813 personnes contactées ayant accepté de répondre au début du questionnaire aux 371 483 personnes de l'échantillon opérationnel, le taux de réponse à l'enquête s'élève à 31,2 %. Il était de 29,8% pour l'enquête « Génération 1998 » et 28,6% pour l'enquête « Génération 2001 ».

---

<sup>17</sup> Calcul à partir du fichier historique d'appels sur la base du classement du dernier appel. Les questionnaires filtres passés lors de la phase de pré qualification sont également pris en compte ici.

**Figure 19. Bilan de collecte pour les individus envoyés en production**

Résultats par individu	Allégé		Complet		Total	
		%		%		%
Réponse à l'enquête	28 127	24,3%	37 262	21,7%	65 389	22,7%
Abandon en cours	1 123	1,0%	2 255	1,3%	3 378	1,2%
Hors champ	23 359	20,2%	28 754	16,7%	52 113	18,1%
Refus de l'individu	3 433	3,0%	6 232	3,6%	9 665	3,4%
A déjà répondu	313	0,3%	691	0,4%	1 004	0,3%
Décédé	255	0,2%	481	0,3%	736	0,3%
Réside à l'étranger	1 940	1,7%	2 233	1,3%	4 173	1,5%
Injoignable permanent	13 816	11,9%	26 191	15,2%	40 007	13,9%
Refus d'un tiers	9 888	8,5%	13 342	7,8%	23 230	8,1%
Hors cible	19 482	16,8%	33 283	19,4%	52 765	18,3%
Ne connaît pas la cible	22	0,0%	6	0,0%	28	0,0%
Pas d'interlocuteur	14 091	12,2%	21 174	12,3%	35 265	12,3%
<b>Total</b>	<b>115 849</b>	<b>100,0%</b>	<b>171 904</b>	<b>100,0%</b>	<b>287 753</b>	<b>100,0%</b>

La figure 19 donne des indications plus précises sur le bilan de collecte, selon le type de questionnaire (complet ou allégé). Il est construit à l'aide du fichier historique d'appels qui regroupe l'ensemble des appels réalisés pour chaque individu, avec l'indication de la conclusion de cet appel. Pour certains individus, plusieurs classements peuvent être disponibles quand plusieurs numéros de téléphone ont été successivement testés, chacun ayant pu aboutir à une conclusion différente. Le tableau présenté s'appuie sur un classement hiérarchisé des différents appels effectués pour joindre l'individu. D'autres choix sont possibles, par exemple en ne retenant que le statut du dernier appel effectué chronologiquement. Les chiffres présentés dans la figure 19 peuvent donc légèrement différer selon les choix effectués.

La comparaison des résultats d'appels de l'enquête « Génération 1998 » et ceux de l'enquête « Génération 2004 » ne montre pas de hausse du refus de répondre mais il y a de plus en plus de difficultés à retrouver les individus, malgré des protocoles de recherches et de pré qualification améliorés.

La mise en place d'un nouveau protocole de recherches élargies de coordonnées téléphoniques explique que la part d'individus classés « Hors Cible » soit élevée.

La figure 20 présente le nombre de répondants obtenus pour les différentes extensions d'échantillon.

**Figure 20. Nombre de répondants collectés sur le champ des différentes extensions**

<b>Champs d'extension</b>	<b>Nombre de répondants collectés</b>
Formations Sanitaires et sociales	5364
Association nationale pour la formation automobile	2939
Sortants de doctorats	1662
Formations de l'Environnement	1781
Formation de la Culture	1266
Formations du Sport	2211
Formations du commerce	3600
Diplômés de niveau L3, M2 et licence professionnelle	13 667
Poitou-Charentes	4076
Provence-Alpes-Côte d'Azur	5342
<i>dont université de Nice</i>	<i>1523</i>
Nord-Pas-de-Calais	4668
Lorraine	4640
Basse-Normandie	3135

#### ***4.4.4. Autres enseignements de la collecte***

Pour le questionnaire complet, la part des répondants<sup>18</sup> est de 37,7 %. 37 262 individus y ont répondu. En moyenne, 26 appels ont été nécessaires pour réaliser une enquête. 43 % des enquêtes ont été réalisées sur des téléphones portables ou des numéros spéciaux. La durée de passation d'un questionnaire complet est passée de 36 minutes le premier jour d'enquête à 26 minutes 30 à la fin du terrain, grâce à une meilleure maîtrise du questionnaire par les enquêteurs, et, pour une petite partie, au traitement de fichiers issus de la qualification (individus ayant validé la partie filtre lors de la pré-qualification). La durée moyenne se stabilise aussi en fin de terrain du fait de l'intégration de nouveaux télé-enquêteurs en période d'apprentissage du questionnaire et du Cati.

Pour le questionnaire allégé, la part des répondants est de 44,0 %. 28 127 individus ont répondu au questionnaire allégé. En moyenne, 22 appels ont été nécessaires pour réaliser une enquête. 37 % des enquêtes ont été réalisées sur des téléphones portables ou des numéros spéciaux. La durée moyenne de passation d'un questionnaire allégé est passée de 19 minutes le premier jour d'enquête à 14 minutes 15 à la fin du terrain.

Pour les deux questionnaires, les évaluations initiales de durée moyenne étaient sous-estimées : le calendrier professionnel prend plus de temps en Cati qu'en test papier. De plus, les personnes qui ont réalisé le test papier avaient une très bonne connaissance du questionnaire et étaient, de ce fait, plus rapides en moyenne que les télé-enquêteurs.

La proportion d'appels sur portables et numéros spéciaux était initialement estimée à 30 %. Malgré une forte incitation des télé-enquêteurs auprès des individus à réaliser les enquêtes sur un téléphone fixe, cet objectif n'a pas pu être atteint. La part des téléphones portables et numéros spéciaux ne cesse

---

<sup>18</sup> Individus acceptant de répondre (enquêtes réalisées et individus hors champ) rapportés aux individus envoyés en production.

d'augmenter au fil des années : 13 % des numéros de téléphone des enquêtes réalisées pour l'enquête « Génération 1998 » (réalisée en 2001) ; 28 % pour l'enquête « Génération 2001 » (réalisée en 2004) ; 40 % pour l'enquête « Génération 2004 » (réalisée en 2007). En particulier, les numéros spéciaux soumis à une facturation plus forte que les portables chez les opérateurs (numéros internet freebox, livebox,...) ont représenté 2,6 % pour l'enquête « Génération 2004 » et risquent d'être encore plus nombreux lors des prochaines enquêtes.

Point positif en revanche, cette évolution montre que l'impossibilité de récupérer les numéros de portables et les numéros spéciaux lors de la phase d'enrichissement en coordonnées téléphoniques ne place pas hors champ les personnes concernées, ces numéros pouvant être récupérés dans les fichiers des établissements ou auprès de tiers lors de la phase de pré-qualification.

## **5. LES TRAITEMENTS EN AVAL (APUREMENT, CODAGE, REpondÉRATION)**

---

A l'issue de la collecte, le prestataire a livré cinq fichiers bruts au Céreq : deux fichiers se rapportent aux questionnaires allégés (calendrier professionnel, reste du questionnaire) ; les trois autres aux questionnaires complets (calendrier professionnel, calendrier du mode de cohabitation et reste du questionnaire).

Différents traitements ont été réalisés sur ces fichiers pour aboutir aux fichiers finaux : apurement des données, enrichissement des résultats de l'enquête par des informations de la base de sondage, traitement de la non réponse globale par repondération, calage sur marge, décomposition de la table du calendrier professionnel en deux tables séparant les périodes d'emploi et les périodes de non emploi, codage du secteur d'activité et de la catégorie socioprofessionnelle, anonymisation des données.

### **5.1. Les tables finales**

A l'issue des différents traitements, les fichiers de résultats de l'enquête « Génération 2004 à trois ans » se répartissent en six tables (hors cas des extensions interrogeant les post-initiaux).

#### ***5.1.1. Les trois tables associées aux questionnaires complets***

Ces tables concernent les personnes du champ Céreq qui ont répondu au questionnaire complet. Les post-initiaux enquêtés pour certaines extensions avec le questionnaire complet ne sont donc pas pris en compte. Permettant de disposer d'informations plus nombreuses, ces tables sont à utiliser pour les exploitations nationales de l'enquête.

**Figure 21. Présentation des fichiers associés au questionnaire complet mis à disposition par le réseau Quételet**

**L'utilisation de la variable de pondération finale (PONDEF) permet de disposer de résultats représentatifs de l'ensemble du champ Céreq.**

Nom	L'observation dans la table est :	Contenu
Table « Individus »	Le répondant (33 655 observations)	Ensemble des réponses de l'enquêté, hors calendrier professionnel mais y compris le calendrier du mode de cohabitation. Informations issues de la base de sondage. Pondérations.
Table « Séquences d'emploi »	La période d'emploi (69 982 observations)	Ensemble des périodes d'emploi présentes dans le calendrier professionnel et variables descriptives de ces périodes. Un identifiant (IDENT) permet d'associer les périodes au répondant concerné de la table « Individus »
Table « Séquences de non emploi »	La période de non emploi (43 381 observations)	Ensemble des périodes de non emploi présentes dans le calendrier professionnel et variables descriptives de ces périodes. Un identifiant (IDENT) permet d'associer les périodes au répondant concerné de la table « Individus »

### *5.1.2. Les trois tables associées aux questionnaires allégés*

Ces tables concernent à la fois les personnes du champ Céreq qui ont répondu au questionnaire allégé et celles qui ont répondu au questionnaire complet. Les post-initiaux enquêtés pour certaines extensions avec le questionnaire complet ou allégé ne sont donc pas pris en compte. Pour les personnes ayant répondu au questionnaire complet, seules les variables disponibles avec le questionnaire allégé ont été conservées. Permettant de disposer d'informations moins nombreuses, ces tables sont à utiliser pour les exploitations associées aux extensions de l'enquête ou, plus généralement, des analyses simples à des échelons fins.

**Figure 22. Présentation des fichiers associés au questionnaire allégé mis à disposition par le réseau Quételet**

**L'utilisation de la variable de pondération finale (PONDEF) permet de disposer de résultats représentatifs de l'ensemble du champ Céreq.**

Nom	L'observation dans la table est :	Contenu
Table « Individus »	Le répondant (59 433 observations)	Ensemble des réponses de l'enquêté, hors calendrier professionnel. Informations issues de la base de sondage. Pondérations.
Table « Séquences d'emploi »	La période d'emploi (125 350 observations)	Ensemble des périodes d'emploi présentes dans le calendrier professionnel et variables descriptives de ces périodes*. Un identifiant (IDENT) permet d'associer les périodes au répondant concerné de la table « Individus »
Table « Séquences de non emploi »	La période de non emploi (78 133 observations)	Ensemble des périodes de non emploi présentes dans le calendrier professionnel et variables descriptives de ces périodes*. Un identifiant (IDENT) permet d'associer les périodes au répondant concerné de la table « Individus »

\* seuls le premier emploi (s'il existe) et la situation à la date de l'enquête sont décrits.

### ***5.1.3. Les trois tables associées à chaque extension de champ***

Deux extensions de champ intégrant les post-initiaux et levant le seuil d'âge de 35 ans ont été réalisées lors de l'enquête « Générations 2004 à trois ans ». L'une, relative aux formations sanitaires et sociales, pour la Drees ; l'autre relative aux formations du sport, pour le ministère de la jeunesse, des sports et de la vie associative.

Chacune de ces deux extensions a donné lieu à la production des trois tables « Individus », « Séquences d'emploi » et « Séquences de non emploi », sur l'ensemble des individus de leurs champs respectifs, qu'ils soient primo sortants ou post-initiaux. Chacune dispose également d'une variable de pondération spécifique.

Les tables contiennent les variables spécifiques au module « post-initiaux » et, pour l'extension des formations du sport, les réponses à l'extension de questionnement.

## **5.2. Des fichiers bruts aux fichiers finaux**

Cette partie présente l'ensemble des traitements réalisés pour construire les fichiers finaux à partir des fichiers bruts, à l'exception du calcul de la pondération finale présentée dans la partie suivante, qui aborde le traitement de la non réponse globale et le calage sur marge.

Les fichiers des questionnaires complets et ceux des questionnaires allégés ont été traités indépendamment mais de façon analogue. Les observations entrant dans le champ Céreq et celles relevant d'extensions spécifiques (les post-initiaux des formations du sport et des formations sanitaires et sociales) n'ont pas été séparées pour les phases d'apurement, de codification, de création de variables. Elles n'ont été dissociées qu'à la fin, pour la constitution des bases finales et les calculs de pondération.

Les traitements effectués sur les bases correspondant aux questionnaires complets seront présentés de façon privilégiée.

### ***5.2.1. La construction des bases brutes***

Rappelons que le prestataire a livré trois fichiers pour les répondants au questionnaire complet : le calendrier professionnel, le calendrier du mode de cohabitation et le reste du questionnaire.

Pour la construction des bases « brutes », les informations du calendrier professionnel ont été éclatées en deux parties de façon à construire une table sur les périodes d'emploi et une table sur les périodes de non emploi. Une table sur les individus a par ailleurs été construite pour rassembler les informations du calendrier de mode de cohabitation et celles du reste du questionnaire, ainsi que des variables synthétiques élaborées à partir du calendrier professionnel.

La table « Individus » contient donc les informations sur l'identification de l'enquêté, son parcours scolaire, ses caractéristiques individuelles et son environnement familial ainsi que les informations demandées dans les modules spécifiques. La table « Séquences d'emploi » décrit chaque séquence

d'emploi à partir du calendrier professionnel. La table « Séquences de non emploi » décrit chaque séquence de non emploi ou de reprise d'études à partir du calendrier professionnel.

Une première phase de validation consiste ensuite à vérifier que l'ensemble des variables présentes dans les questionnaires étaient remplies et, pour chaque variable, que les modalités de réponses correspondaient à celles des questionnaires.

Dans la table « Individus », des variables relatives à la trajectoire, aux diplômes, au calendrier professionnel et au calendrier habitat ont été créées.

#### **Encadré 7 : Les variables de la table « Individus » construites à partir des calendriers**

##### Variables issues du calendrier professionnel

Les variables MOIS1 à MOIS45 permettent de suivre mois par mois, de novembre 2003 (mois1) à la date d'interrogation, entre avril 2007 et juillet 2007 (mois 45), la situation professionnelle du répondant (elles fournissent également l'identifiant de la période concernée dans la table des périodes d'emploi ou celle des périodes de non emploi).

Les variables ID, IF et DUROBS correspondent respectivement aux numéros du premier et du dernier mois observé et au nombre total de mois renseignés dans le calendrier.

Les variables NMEMP, NMJVAC, NMCHO, NMINA, NMFOR, NMETU et NMVAC renseignent sur le nombre total de mois passés dans chaque situation professionnelle, respectivement en emploi, en petit boulot d'été, au chômage, en inactivité, en formation, en études et en vacances<sup>19</sup>.

Les variables NSEMP, NSJVAC, NSCHO, NSINA, NSFOR, NSETU et NSVAC renseignent sur le nombre total de séquences passées dans chaque situation professionnelle. La variable NSTOT donne le nombre total de séquences de l'individu.

La variable TAPE renseigne sur le temps d'accès au premier emploi, TRAJPRO sur l'enchaînement des séquences et SITDE sur la situation professionnelle à la date d'enquête.

##### Variables issues du calendrier du mode de cohabitation

Pour les répondants au questionnaire complet, les variables HMOIS1 à HMOIS45 ont été créées pour donner le mode de cohabitation mois par mois de novembre 2003 (mois1) à la date d'interrogation, entre avril 2007 et juillet 2007 (mois 45).

---

<sup>19</sup> Les « petits boulots d'été » désignent des périodes d'emploi terminées au plus tard en septembre 2004 et déclarées comme emplois de vacance par l'enquêté. Les « vacances » sont l'une des modalités possibles pour caractériser la première séquence du calendrier, juste à l'issue de la formation initiale, quand sa durée est inférieure ou égale à 3 mois. Dans ce cas, l'enquêté peut déclarer cette période comme une période de « vacances ».

### **Encadré 8 – Les réponses à choix multiples**

Certaines questions autorisent les réponses multiples. Après codage des réponses ouvertes, le cas échéant, les différentes réponses possibles ont donné lieu à la création d'autant de variables dichotomiques, chaque modalité de réponse proposée étant associée à une variable binaire (Oui/Non).

Pour le questionnaire complet, les questions concernées sont les questions Q25, Q35, Q39A, Q39D, Q39F, Q39H, Q39H2, Q39H3, TH4/O\_TH4, CA26/O\_CA26 (table « Individus ») ; les questions EX4, EX5 (table « Séquences d'emploi ») et les questions RP2, RP2A, RA8, RP13, RA18 (table « Séquences de non emploi »).

Pour le questionnaire allégé, les questions concernées sont les questions RP2, RP2A, RA8, RP13 et RA18 (table « Séquences de non emploi »).

#### ***5.2.2.La vérification des filtres***

Le respect des filtres a été contrôlé. Selon leurs caractéristiques et leurs réponses aux questions précédentes, les individus ne doivent pas tous répondre aux mêmes items. En règle générale, les filtres sont gérés par le CATI, sauf dans le cas où l'enquêteur procède à des retours arrière lors de la passation du questionnaire. Dans ce cas, certaines variables qui n'auraient pas dû être renseignées le sont. Les contrôles de cohérence ont conduit à les mettre à blanc a posteriori (quelques dizaines de cas de correction au plus sur certaines questions).

#### ***5.2.3.La suppression d'observations non exploitables***

En théorie, un calendrier d'activité mal renseigné ne pouvait être validé et l'enquête ne pouvait alors pas se poursuivre. En pratique, il est arrivé que, dans certains cas, plusieurs calendriers professionnels soient simultanément ouverts dans l'application de collecte. De ce fait, certains calendriers d'activité ne sont pas exploitables, parce que certaines périodes se chevauchent, ne sont pas renseignées ou parce que des périodes d'emploi mentionnées ne sont pas associées à une description de l'emploi alors que cela aurait dû être le cas. Ces observations jugées non exploitables sur le noyau du questionnaire ont été supprimées.

Par ailleurs, certains répondants se sont révélés hors champ a posteriori. Quelques doublons ont également été supprimés.

Au total moins de 3 % des observations récupérées ont été supprimées.

#### **5.2.4. Le traitement des salaires**

Les salaires déclarés par les individus sont contrôlés. Pour cela, ils sont transformés en équivalent temps plein pour être comparés à une norme, calculée à partir du taux horaire du SMIC et des taux réglementés. Lorsque le salaire s'avère inférieur à cette norme, il est imputé à l'aide d'une équation de salaire ou de la médiane calculée sur une population homogène. Les valeurs extrêmes sont ensuite étudiées par niveau de diplôme et une analyse fine permet de détecter et corriger le cas échéant les valeurs aberrantes.

Il convient de souligner que cette phase est une étape délicate. Outre le fait de s'appuyer sur des règles *ad hoc*, celles-ci doivent parfois s'appliquer à des informations imparfaites. Par exemple, les situations de travail à temps partiel mal déclarées sont difficiles à traiter. Les réponses brutes sont donc proposées dans les fichiers au côté de l'information redressée.

#### **5.2.5. L'enrichissement par des informations de la base de sondage**

La table des « Individus » est enrichie par des informations issues de la base de sondage et relatives à la scolarité suivie en 2003-2004 : formation suivie, niveau, spécialité, ...

Ces informations complètent les réponses au questionnaire sur les informations concernant le parcours scolaire, abordé de façon très détaillée dans le questionnaire complet de l'enquête<sup>20</sup>.

Ensemble, elles permettent de définir le diplôme de sortie et de reconstituer l'ensemble des diplômes obtenus par les répondants au questionnaire complet<sup>21</sup>. Pour les répondants au questionnaire allégé, l'ensemble des diplômes obtenus n'est pas disponible mais le plus haut diplôme obtenu est disponible<sup>22</sup>.

#### **5.2.6. Le repérage du plus haut diplôme obtenu**

Pour les répondants au questionnaire complet, le plus haut diplôme est créé à partir d'une hiérarchisation normalisée de l'ensemble des diplômes décrits dans le parcours scolaire. Cette normalisation est *ad hoc*.

Pour le questionnaire allégé, la notion de plus haut diplôme est laissée à l'appréciation du répondant lorsqu'on lui demande s'il « possédait un diplôme supérieur ou équivalent » à son diplôme de sortie en 2004, puis de préciser quel était ce « plus haut diplôme ».

---

<sup>20</sup> Pour construire les variables de diplômes, les variables suivantes de la base de sondage sont mobilisées : DIPL, STRATE, STRATEGN et variables issues du fichier Sise pour l'Université. Sont également utilisées les déclarations des individus à la question Q7B et au module parcours scolaire (questions Q31 à Q42 pour le questionnaire complet et questions Q34 à Q39E pour le questionnaire allégé).

<sup>21</sup> Pour le questionnaire complet, les variables NIVEAU, NIVSOR9, CLASSE et NSFSOR sont créées pour le diplôme de sortie ; les variables PHDIP, PHDIP9, NSFPHD, pour le plus haut diplôme obtenu ; les variables DIPC1 à DIPC11, NSFD1 à NSFD11, pour l'ensemble des diplômes du parcours.

<sup>22</sup> Pour le questionnaire allégé, les variables PHDIPREG, Q39AJ2 et Q39E2 sont créées pour le plus haut diplôme obtenu, ainsi que CLASSE et NIVEAU, pour le diplôme de sortie.

Le plus haut diplôme obtenu ne correspond pas toujours au niveau de formation de la classe de sortie, par exemple du fait des échecs aux examens et abandons en cours d'études. La table de passage de l'un à l'autre obtenue avec le questionnaire complet n'est pas la même que celle obtenue avec le questionnaire allégé, sans qu'il soit possible de distinguer ce qui relève de l'effet d'échantillonnage de ce qui résulte de la différence de questionnement, les questions posées sur le plus haut diplôme dans le questionnaire allégé n'étant pas posées dans le questionnaire complet.

### ***5.2.7. Le codage des spécialités de formation***

Les spécialités de formation ont été codées selon la Nomenclature des spécialités de formation (NSF), dans sa version définie par le décret n° 94-522 du 21 juin 1994 portant approbation de la Nomenclature des spécialités de formation<sup>23</sup>.

Cette nomenclature, élaborée au sein du Conseil national de l'information statistique, sert de référence pour les textes officiels, les travaux d'études et les systèmes d'information des administrations et des établissements d'éducation (articles D311-1 à D311-4 du code de l'Education).

Elle comporte trois niveaux emboîtés : les deux premiers niveaux (4 postes et 17 postes) fixent la liste des « domaines de spécialités » en matière de formation. Le troisième niveau (93 postes) fixe la liste des « groupes de spécialités » de formation.

A partir des informations présentes dans les variables des fichiers Sise, une table de passage a été créée par la DEPP pour coder le diplôme en NSF.

### ***5.2.8. Le codage du secteur d'activité des périodes d'emploi***

Pour les périodes d'emploi, le secteur d'activité des entreprises a été codé selon la Nomenclature économique de synthèse (NES), dans sa version datant de 1994.

Cette nomenclature a été créée par l'Insee pour disposer d'une classification pertinente pour l'analyse économique, en s'appuyant autant que possible sur les comportements d'agents confrontés à leur marché, là où les nomenclatures officielles d'activité, comme la NAF, font intervenir d'autres critères comme les spécificités techniques du processus de production ou l'organisation en filières de production.

La NES comprend trois niveaux emboîtés comportant respectivement 16, 36 et 114 positions. Pour chaque séquence en entreprise décrite par l'individu, l'activité économique de l'entreprise a été codée à partir de la nomenclature économique de synthèse (NES) en 114 postes et selon un regroupement propre au Céreq en 67 postes (voir annexe 5).

L'activité économique de l'entreprise a été codée à partir des questions EP2A à EP10 sur le service ou produit principalement fabriqué ou vendu, selon des règles *ad hoc* fixées par le Céreq. Trois méthodes de codage ont été utilisées séquentiellement :

---

<sup>23</sup> Les variables créées selon cette nomenclature sont NSFSOR, NSFD1 à NSFD11, NSFPHD.

- pour certaines entreprises, listées dans un menu déroulant du logiciel de collecte, un codage direct était réalisé lors de la passation du questionnaire, à partir d'une table de passage *ad hoc* (26 680 séquences pour le questionnaire complet et pour le questionnaire allégé).
- un codage automatique à partir du nom de l'entreprise et des questions sur le service ou produit principalement fabriqué ou vendu (de l'ordre de 20 000 séquences pour le questionnaire complet et 80 000 séquences pour le questionnaire allégé).
- un codage manuel par une société extérieure pour les séquences n'ayant pas pu être codées par les deux méthodes précédentes (21 806 séquences pour le questionnaire complet et 16 342 pour l'allégé).

### ***5.2.9. Le codage de la profession des périodes d'emploi***

Pour les répondants au questionnaire complet, les situations professionnelles de début et fin de séquence d'emploi sont détaillées. Pour les répondants au questionnaire allégé, seules certaines situations sont détaillées (situation à l'embauche pour le premier emploi, situation à la date de l'enquête pour l'emploi actuel, situation à l'embauche et à la date de l'enquête si un seul emploi qui dure encore à la date de l'enquête). L'explication donnée ci-dessous s'applique au questionnaire complet ; les questions sont les mêmes pour le questionnaire allégé.

La profession a été codée selon la nomenclature des professions et catégories socioprofessionnelles (PCS) dans sa version datant de 2003<sup>24</sup>.

Cette nomenclature a été créée par l'Insee pour disposer d'une classification pertinente pour l'analyse des comportements socio-économiques des personnes. Elle classe la population selon une synthèse de la profession (ou de l'ancienne profession), de la position hiérarchique et du statut (salarié ou non).

La PCS comporte trois niveaux d'agrégation emboîtés : 8 groupes socioprofessionnels, 24 ou 42 catégories socioprofessionnelles et 486 professions.

La construction de la PCS s'appuie d'abord sur le libellé de l'emploi demandé à l'individu<sup>25</sup>, complété par d'autres informations relatives à la position professionnelle de la personne<sup>26</sup> et aux caractéristiques de l'entreprise dans laquelle elle travaille (appartenance au secteur public, taille de l'entreprise, secteur d'activité<sup>27</sup>). Des corrections orthographiques ont été effectuées sur une partie des libellés, la plupart étant cependant issue d'un menu déroulant très détaillé proposé par l'application de collecte.

---

<sup>24</sup> Dans la version 2003 de la PCS, les premier et deuxième niveaux sont restés inchangés par rapport à la version de 1982. La rénovation de 2003 n'a porté que sur le niveau le plus fin (regroupement de certaines professions dont la distinction était devenue obsolète et éclatement d'autres afin de tenir compte de l'apparition de nouveaux métiers ou de nouvelles fonctions).

<sup>25</sup> Variables EMPL\_EMB et EMPL\_FIN.

<sup>26</sup> Variables POSPRO\_EMB, POSPRO\_FIN, construites à partir des questions EP33 à EP36B, apurées au préalable.

<sup>27</sup> L'appartenance au secteur public (variable PUB) est déterminée en utilisant une liste d'entreprises, d'une part ; et la question EP3, d'autre part. La taille de l'entreprise est uniquement déclarative (question EP12). Le secteur d'activité est défini à partir de la question EP38 et du code NAF reconstitué à partir de la NES.

A partir de ces informations, le logiciel SICORE, développé par l'INSEE, permet de coder automatiquement la PCS. Les professions qui n'ont pas pu être codées par cette procédure ont été transmises à une société externe qui a procédé à une codification manuelle.

Pour les répondants au questionnaire complet, 72 707 professions ont été codées : 50 987 automatiquement et 21 720 en manuel. Pour les répondants au questionnaire allégé, 41 493 professions ont été codées : 25 875 automatiquement et 15 618 en manuel.

### **5.2.10. Le codage du contrat de travail**

Pour les périodes d'emploi, plusieurs questions abordent la nature du contrat de travail en début de séquence et en fin de séquence, permettant par exemple de repérer les personnes qui accèdent à une entreprise via un emploi court (contrat à durée déterminée de droit commun, emploi aidé, mission d'intérim) puis y restent après passage en contrat à durée indéterminée<sup>28</sup>.

Deux variables de synthèse sont donc construites pour définir le contrat de travail en début de séquence (stat\_emb) et le contrat de travail en fin de séquence (stat\_fin). La figure 23 présente les différentes modalités existantes et les questions mobilisées pour les constituer<sup>29</sup>.

---

<sup>28</sup> Une seule période d'emploi est constituée en cas de changement de contrat au sein de la même entreprise tant qu'il n'y a pas interruption de la présence du salarié dans l'entreprise, ni changement d'établissement. Pour les personnes qui reviennent dans une entreprise dans laquelle elles sont déjà passées après une interruption d'au moins un mois, deux séquences distinctes sont décrites. Une variable a toutefois été créée pour permettre de repérer ces cas : la variable REFSEQ. Si l'individu n'a pas travaillé dans l'entreprise auparavant, REFSEQ = NSEQ (numéro identifiant de la séquence). Si l'individu a déjà travaillé dans l'entreprise alors REFSEQ est égale au numéro de la première séquence qu'il a effectuée dans cette entreprise.

<sup>29</sup> Au total, les informations suivantes sont utilisées : variable INTER (intérimaire plusieurs missions) et réponses aux questions EP1, EP20, EP21, O\_EP21, EP24, EP25, EP 26, EP 27, O\_EP 27.

**Figure 23. Codification du contrat de travail**

<b>Modalité</b>	<b>Nature du contrat</b>	<b>Questions mobilisées</b>
01.	Indépendant (à son compte)	<i>EP1</i>
02.	Aide familial	<i>EP1</i>
03.	Fonctionnaire	<i>EP20 /EP26</i>
04.	CDI	<i>EP20 /EP26</i>
05.	CDD	<i>EP20 /EP26</i>
06.	Apprentissage	<i>EP21 /EP27</i>
07.	Intérimaire	<i>INTER, EP20 /EP26</i>
08.	Contrat de professionnalisation (ou de qualification)	<i>EP21 /EP27</i>
10.	CAE/CES/CEC	<i>EP21 /EP27</i>
11.	Autres contrats aidés	<i>EP21 /EP27</i>
12.	Saisonnier	<i>EP21 /EP27</i>
13.	Stagiaire	<i>recodage de O_EP21/O_EP27</i>
14.	Pas de contrat	<i>recodage de O_EP21/O_EP27</i>
16.	Autres	<i>EP21/EP27</i>
18.	CIE (contrat initiative emploi)	<i>EP21/EP27</i>
20.	ATER	<i>recodage de O_EP21/O_EP27</i>
21.	Post doctorat	<i>recodage de O_EP21/O_EP27</i>
22.	Vacataire	<i>recodage de O_EP21/O_EP27</i>
23.	Contrat étranger	<i>recodage de O_EP21/O_EP27</i>
25.	CNE	<i>EP20 /EP26</i>

### **5.2.11. Les variables de localisation géographique**

Pour l'ensemble des répondants (questionnaire complet ou allégé), plusieurs localisations géographiques sont disponibles : la commune de résidence en 6<sup>ème</sup>, au baccalauréat (le cas échéant) et à la date de l'enquête ; ainsi que le département de l'établissement de formation de l'année scolaire 2003-2004. Pour chaque séquence en entreprise, la commune du lieu de travail est également renseignée.

Les communes sont repérées à partir de leur code officiel géographique, attribué par l'Insee, ce qui permet de reconstituer ensuite différentes codifications géographiques : région et département, zone d'emploi et zonage en aire urbaine, deux zonages d'études élaborés par l'Insee<sup>30</sup>.

<sup>30</sup> Dans les tables « Individus », les variables ainsi créées sont les suivantes : AU6EME, DEP6EME, REG6EME, ZE6EME, POL6EME (à partir de la commune de résidence en 6<sup>ème</sup>), AUBAC, DEPBAC, REGBAC, ZEBAC, POLBAC (à partir de la commune de résidence au bac), AUINTER, DEPINTER, REGINTER, ZEINTER, POLINTER (à partir de la commune de résidence à la date de l'enquête), REGETAB (à partir du département de l'établissement de formation). Dans les tables « Séquences d'emploi », il s'agit des variables AU, DEP, REG, ZE, POL, créées à partir du code commune de l'entreprise.

Deux informations sont issues du zonage en aires urbaines : le type d'espace auquel la commune appartient (pôle urbain, espace périurbain monopolarisé, espace périurbain multi polarisé, ou espace à dominante rurale) et, pour les communes appartenant à un pôle urbain ou à un espace périurbain monopolarisé, le code de l'aire urbaine concernée.

#### 5.2.12. *Le codage des réponses en clair*

Certaines variables présentent une modalité « Autre », qui ouvre la possibilité d'une réponse en clair. Dans la mesure du possible, les réponses en clair données par les individus ont été recodées.

Pour le questionnaire complet, dix-neuf variables ont fait l'objet de recodages.<sup>31</sup> Pour le questionnaire allégé, ce sont quatorze variables qui ont fait l'objet d'un traitement des ouverts<sup>32</sup>.

Toute réponse ouverte donnée par un individu a été analysée. Lorsqu'un reclassement dans une modalité existante était possible, il a été effectué. Lorsque des modalités non proposées dans le questionnaire étaient citées de manière récurrente dans les ouverts, ces nouvelles modalités ont été créées. Les réponses n'ayant pas pu être reclassées selon les précédents critères sont restées dans la modalité « Autre ».

#### 5.2.13. *Le contrôle de cohérence finale*

A l'issue des apurements et des codages détaillés dans les paragraphes précédents, un nouveau contrôle de cohérence a été réalisé, notamment pour vérifier à nouveau les filtres. La cohérence entre salaire, temps de travail, profession a également été vérifiée, pour corriger à nouveau, le cas échéant, le salaire.

**L'ensemble des contrôles de cohérence et des retraitements, comme les pondérations, a été effectué en vue des utilisations standards des fichiers d'études, notamment dans la perspective de la publication des premiers résultats dans la série des *Bref* puis de l'ouvrage de résultats détaillés *Quand l'école est finie*. La réalisation de travaux très détaillés sur des publics spécifiques ou sur des problématiques ciblées peut nécessiter de réaliser des contrôles complémentaires et, le cas échéant, des corrections spécifiques.**

---

<sup>31</sup> Table « Individus » : Q34, Q34B, Q36, Q36B, Q39 (parcours scolaire), TH3, TH6 (module thèse), SP16, SP24 (module post-initiaux), CA2, CA4, CA6, CA23D, CA26 (caractéristiques individuelles). Table « Séquences d'emploi » : EP21 et EP27 (contrats de travail). Table « Séquences de non emploi » : ETUP1, ETUA2 et ETUA4 (reprise d'études).

<sup>32</sup> Table « Individus » : Q36, Q36B, SP16, SP24A, SP24B, SP26, SP261, CA4 et CA6. Table « Séquences d'emploi » : EP21 et EP27. Table « Séquences de non emploi » : ETUP1, ETUA2, ETUA4.

#### **5.2.14. L'anonymisation des fichiers**

Afin d'assurer la confidentialité des réponses des personnes enquêtées et conformément au principe du secret statistique garanti par la loi de n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, les fichiers de résultats ont été rendus anonymes.

L'anonymisation consiste d'abord à supprimer les informations directement nominatives (nom et prénom des personnes), ainsi que les coordonnées téléphoniques et postales, nécessaires à la réalisation de la collecte de l'enquête. Les noms des entreprises sont également supprimés.

Elle consiste aussi à s'assurer qu'une identification indirecte par recoupement des différentes informations disponibles est impossible. Pour cela, certaines variables ont été supprimées, d'autres ont vu leurs modalités regroupées. En particulier, les informations de localisations géographiques fines ne sont pas diffusées (code commune, code des aires urbaines). Dans les questions ouvertes, les réponses sont examinées pour supprimer toute information de nature à faciliter l'identification de la personne (mention d'un nom d'entreprise, d'un lieu précis, etc.).

### **5.3. La pondération finale**

La pondération finale combine plusieurs facteurs : le taux d'échantillonnage, la correction du défaut de couverture de la base, la correction de la non réponse totale et un calage sur marges. Le taux d'échantillonnage résultant lui-même d'un coefficient de dilatation lié à la couverture estimée, d'un coefficient de dilatation lié à la non-réponse attendue et d'un coefficient d'échantillonnage moyen.

Plusieurs pondérations ont été construites :

- une pour exploiter les questionnaires complets des individus du champ Céreq, soit 33 655 individus ;
- une pour exploiter l'ensemble des questionnaires allégés des individus du champ Céreq (en y incluant les questionnaires complets mis au format de l'allégé), soit 59 433 individus ;
- une pondération spécifique a été réalisée pour les individus dans le champ de la DREES (donc y compris les individus post initiaux hors du champ Céreq) ;
- une pondération spécifique a été réalisée pour les individus dans le champ du Sport (donc y compris les individus post initiaux hors du champ Céreq) ;

La pondération PONDEF fournie dans les bases résulte de la multiplication de plusieurs coefficients différents. Seules sont décrites ci-dessous les grandes étapes de la pondération ; nous ne présentons pas les cas particuliers (post-initiaux et coefficients négligeables).

La pondération a été effectuée selon le même principe pour les fichiers associés au questionnaire complet et pour ceux associés au questionnaire allégé. Les chiffres présentés ci-dessous sont relatifs à la pondération du complet. La figure 24 compare la dispersion des poids des différentes enquêtes successives.

**Figure 24. Dispersion des pondérations dans les enquêtes « Génération » à trois ans**

Génération	Moyenne	Écart-type	Minimum	Maximum
1998	13,4	13,45	1	721
2001	54,4	46,25	2	448
<b>2004</b>	<b>21,9</b>	<b>17,77</b>	<b>1</b>	<b>264</b>

La pondération finale PONDEF s'obtient par multiplication de six coefficients :

$$PONDEF = COEFF1 \times COEFF2 \times COEFF3 \times COEFF4 \times COEFF5 \times COEFF6$$

Avec :

COEFF1 (Moyenne = 1,32)

Ce coefficient vise à redresser les inégalités de couverture de la base de sondage. Le degré de finesse de ce coefficient est variable selon les champs.

COEFF2 (Moyenne = 5,24)

Ce coefficient correspond directement à l'inverse du taux de tirage. Il tient notamment compte de l'utilisation ou non de l'échantillon de réserve. Si la réserve est utilisée : coeff2 = 1 / taux de tirage. Sinon, coeff2 = (1 / taux de tirage)\*1,1

COEFF3 (Moyenne = 1,01)

Après le tirage de l'échantillon, certains individus ont pu être repérés comme étant en poursuite d'étude. Ces individus ont donc été isolés et retirés de l'échantillon. Ce troisième coefficient vise à corriger l'impact de cet apurement postérieur au tirage. Il concerne principalement des individus poursuivant en IUFM et CFA.

COEFF4 (Moyenne = 1,26)

Ce coefficient cherche à corriger la non-réponse liée à l'absence de coordonnées téléphoniques, les personnes concernées n'ayant pas été envoyées en plateau d'enquête.

COEFF5 (Moyenne = 2,56)

Ce coefficient cherche à corriger la non-réponse des individus envoyés en plateau d'enquête, associée aux refus, aux injoignables permanents et aux hors cibles, les répondants à l'enquête (37 000 pour les questionnaires complets) et les individus répondants et classés hors champs lors des questions de filtrage, sont considérés comme « répondants ».

COEFF6 (Moyenne = 1,08)

Ce sixième et dernier coefficient est un coefficient de recalage sur des données issues de l'Enquête Emploi de l'INSEE par grand niveau de formation.

### *5.3.1. La prise en compte de la non réponse globale*

Pour le calcul des coefficients 4 et 5, des modèles de réponse ont été mis en œuvre. Plusieurs modèles logistiques ont été testés pour choisir les variables à retenir et définir le modèle de réponse à privilégier ; d'une part, pour modéliser la probabilité de ne pas avoir de numéro de téléphone ; d'autre part, pour modéliser la probabilité de ne pas répondre au questionnaire lorsqu'un numéro de téléphone est disponible.

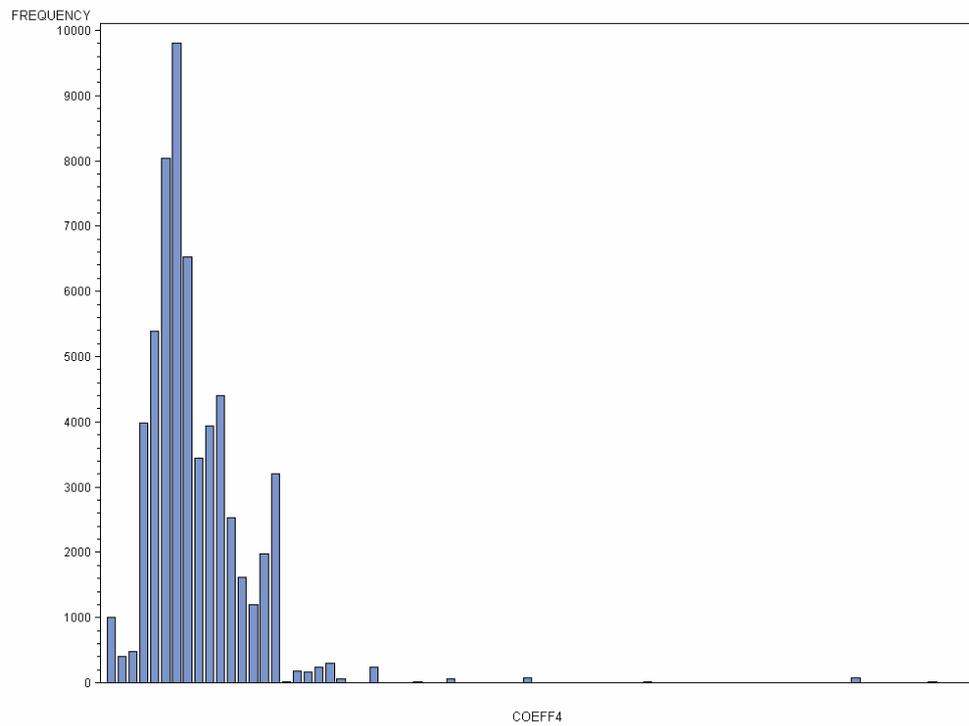
Les différents tests ont mis en évidence le rôle majeur de trois variables discriminantes : la strate de formation, la région de l'établissement de formation et la méthode de récupération du numéro de téléphone.

Les variables discriminantes retenues permettent, par croisement des modalités qu'elles prennent, de définir des regroupements d'individus de même caractéristiques. Pour chaque regroupement un coefficient de redressement est ensuite calculé, chacun des deux coefficients étant calculés séparément.

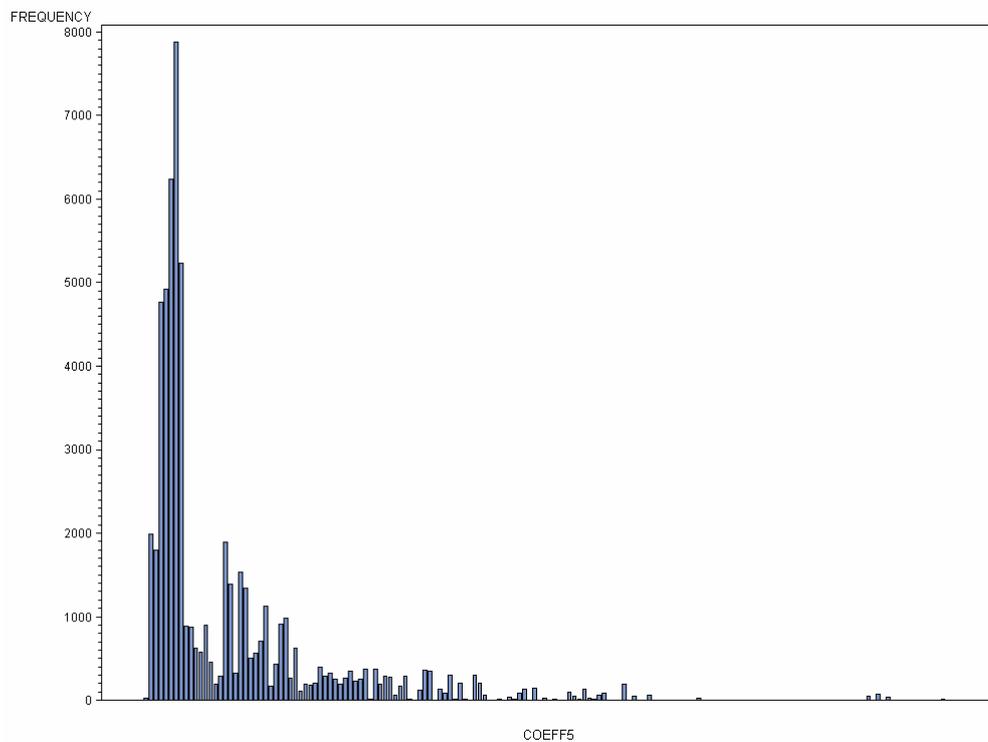
Au préalable, des regroupements de modalités ont été réalisés pour les régions et les strates de formation afin de minimiser la variance des coefficients. Ces regroupements ont été effectués à l'aide de classifications ascendantes hiérarchiques sur les coefficients 4 et 5 issus des modèles logistiques, respectivement pondérés par les effectifs des régions et des strates.

Profitant de la taille de l'échantillon, les travaux ont été conduits séparément pour les sortants issus des fichiers transmis par les rectorats et les autres.

**Figure 25. Distribution du coefficient 4 (correction de la non réponse liée à l'absence de coordonnées téléphoniques)**



**Figure 26. Distribution du coefficient 5 (correction de la non réponse des personnes contactées)**



Variable	N	Moyenne	Ecart-type	Minimum	Maximum
coeff4	59433	1.4386458	0.2682523	1.0395257	8.9090909
coeff5	59433	2.2929285	1.1053569	1.0000000	16.4761905

### 5.3.2. Le calage sur marges

A l'issue du traitement de la non réponse, un calage sur marges a été effectué de façon à retrouver des effectifs totaux de sortants correspondant à ceux estimés avec l'enquête Emploi de l'INSEE, avec leurs pondérations disponibles à la date des traitements, cet effectif total étant ventilé à partir d'une répartition par plus haut diplôme atteint issu de la DEPP. La figure 27 présente les marges utilisées, estimées en compilant les enquêtes trimestrielles de trois années successives, 2004 à 2006.

**Figure 27. Marges employées**

Plus haut diplôme	Effectifs	%
Sans qualification (niveau VI et Vbis)	48 642	6,6
Sans diplôme mais qualifié (non diplômés de niveau V et IV)	74 437	10,1
CAP – BEP et équivalent (diplômés de niveau V)	127 501	17,3
BAC et équivalent (diplômés de niveau IV)	176 880	24
Supérieur court (diplômés de niveau III)	137 082	18,6
Supérieur long (diplômés de niveau II et I)	172 458	23,4
Ensemble	737 000	100

Source : Insee, Enquêtes emploi 2004, 2005, 2006, pondération 2006.

## **6. BILAN ET PERSPECTIVES POUR LES ENQUÊTES FUTURES**

---

Pour préparer et mettre en œuvre l'enquête Génération 2004, quatre années ont été nécessaires (entre 2005 et 2008). La sollicitation des établissements de formation a permis de constituer une base de sondage d'environ 1 250 000 lignes. L'échantillon tiré contenait pour sa part quelques 400 000, dont un peu moins de 300 000 ont été envoyés sur un plateau de pré-qualification ou d'enquête. Un total de 65 000 questionnaires a finalement été collecté.

L'introduction d'une nouvelle phase de recherche téléphonique a permis de récupérer davantage de sortants échantillonnés et d'améliorer ainsi le taux de succès de l'enquête, malgré le choix de réduire le nombre maximum de numéros de téléphones récupérables lors de la phase d'enrichissement. De ce point de vue, le choix de privilégier une hiérarchisation des numéros et d'élever le nombre de rappels possibles pour un même numéro s'est avéré efficace.

La passation du questionnaire allégé s'est bien déroulée. Elle a permis de limiter les coûts. Un bilan global des exploitations réalisées sera sans doute nécessaire pour le confirmer mais les premières publications réalisées montrent que ce questionnaire court permet de produire à des niveaux de détail fin l'essentiel des indicateurs standards.

En revanche, la gestion de l'enquête s'est révélée complexe. D'une part, la structure des questionnaires et la forte personnalisation des questions par la mise en place de filtres ont complexifié le développement des applications de collecte par téléphone. D'autre part, la mise en œuvre de deux questionnaires s'est traduite par une charge de travail proche de celle de deux enquêtes simultanées : développement et test de deux questionnaires, suivi et validation de deux types de fichiers, création de deux jeux de bases indépendants, y compris sur les apurements et recodages, élaboration de deux pondérations, création d'une table de passage entre les bases complètes et allégées, cohérence des bases à assurer, ....

Concernant les coûts, le temps de passation des questionnaires a été sous-estimé, en particulier pour le questionnaire allégé. Initialement prévu à 12 mn, il a finalement eu une durée moyenne de 14 mn. Cette sous-estimation du temps peut s'expliquer par une passation du calendrier professionnel plus rapide sous format papier, par exemple dans le cas où l'enquêteur doit corriger des éléments de calendrier déjà saisis. D'autre part, les enquêteurs du Céreq, qui connaissaient parfaitement le questionnaire pour l'avoir testé auparavant, étaient en moyenne plus rapides que ceux du prestataire, d'autant que celui-ci a connu un turn-over important des télé-enquêteurs.

Enfin, la gestion des extensions, particulièrement nombreuses, s'est révélée délicate pour le dimensionnement des plateaux d'enquête au vu du volume total d'enquêtes à réaliser (environ 70 000), mais aussi pour assurer le contrôle des télé-enquêteurs et pour gérer les échantillons de réserve.

Plusieurs pistes d'amélioration pourraient être explorées pour l'avenir du dispositif.

*Pour la constitution de la base de sondage*, le travail en amont et pendant la collecte auprès des établissements des fichiers de sortants est crucial : sélection des établissements dans le champ de l'enquête, choix des variables demandées, contrôles de cohérence, couplage des fichiers et vérification des effectifs. L'effort engagé doit être poursuivi pour obtenir le maximum de numéros de téléphone dans la base de sondage. Une meilleure couverture des numéros de téléphone dans la base de sondage (existence, téléphone fixe ou portable, nombre de téléphones disponibles) améliore le taux de succès à l'enquête et réduit les coûts de recherches téléphoniques.

*Pour la conception du questionnaire d'enquête* : l'utilisation d'un questionnaire unique est à privilégier (complet ou allégé), ainsi qu'une attention particulière pour l'estimation de sa durée. L'intérêt de prévoir la prise en compte d'emplois simultanés dans plusieurs entreprises pourrait être expertisé, de même que la façon de traiter des professions telles que les pigistes ou les intermittents du spectacle dans le calendrier professionnel.

*Pour la recherche de coordonnées téléphoniques* : l'absence d'annuaire sur les téléphones portables et numéros spéciaux est problématique. Il risque de l'être de plus en plus à l'avenir avec la hausse du dégroupage et des téléphones portables comme unique numéro. Les possibilités d'amélioration dépendent ici beaucoup d'évolutions institutionnelles sur lesquelles les concepteurs d'enquêtes ont peu de prise.

*Concernant l'organisation des plateaux d'enquête* : la dissociation temporelle entre la phase de pré-qualification et la phase d'enquête entraîne une certaine déperdition : certains individus joints dans la phase de pré-qualification refusent ensuite d'être recontactés plus tard pour poursuivre le questionnaire au-delà des questions filtres déjà posées ; d'autres se révèlent injoignables ensuite.

*Concernant la formation des télé-enquêteurs* : il convient d'insister sur les différents champs couverts par l'enquête (Céreq et extensions) et sur l'attention à apporter pour le recueil des dates sur lesquelles les enquêtés ont parfois des difficultés de mémoire (dans la partie filtre et pour le parcours scolaire notamment).

Enfin, l'utilisation des nouvelles technologies doit être poursuivie, tant pour la collecte de la base de sondage que pour les contacts établis avec les enquêtés (utilisation du courriel, possibilité d'une saisie en ligne de certaines informations ...).



## **ANNEXE 1 : COMPLÉMENTS SUR LES POIDS DE TIRAGE**

---

Cette annexe complète la partie 3 relative au plan de sondage en présentant les valeurs moyennes par strates agrégées de deux des trois composantes du poids de tirage : le coefficient de dilatation lié à la couverture et le coefficient de dilatation lié à la non réponse attendu - ce dernier tient compte en fait de l'ensemble des facteurs de non succès, y compris l'impossibilité à joindre la personne ou sa mise hors champ, notamment en cas de poursuite d'études.

Le taux de tirage se décompose ainsi :  $Tx = P1 * P2 * P3$  avec :

- $P1$ , le coefficient de dilatation lié à la couverture
- $P2$ , le coefficient de dilatation lié à la non-réponse attendue
- $P3$ , le coefficient d'échantillonnage.

### **Le coefficient $P1$**

Pour le champ « Université », il est calculé sur le croisement degrés-discipline (agrégée en 6 postes)-région.

Pour le champ « Apprentis » (CFA et non CFA), il est calculé sur le croisement spécialité-région-niveau.

Pour les « Ecoles du supérieur », il est calculé par strate (variable Strateg).

Pour les « IUFM », il est calculé pour le croisement région-strate (variable Strateg).

### **Le coefficient $P2$**

Les coefficients  $P2$  sont calculés sur le croisement strate de formation - champ - région. Le degré de finesse de la variable de strate de formation utilisée dépend des effectifs présents dans la strate.

Ce coefficient est égal au nombre de questionnaires à mettre en production pour obtenir un répondant dans le champ.

Les coefficients ont été calculés à partir des observations faites sur l'enquête « Génération 98 » qui compte un échantillon plus important que l'enquête « Génération 2001 », ce qui permet de calculer des coefficients sur des croisements plus fins de variables.

### **Le coefficient $P3$**

Les valeurs moyennes du coefficient  $P3$  ne sont pas présentées ici. La partie 3.4 détaille son mode de calcul. Il vaut au départ  $1/25^e$ , niveau corrigé ensuite selon les individus de façon à assurer les sur-représentations souhaitées dans le cadre des extensions d'échantillon.

***Coefficient P1 moyen par strate (variable Strateg)***

<b>CHAMP</b>	<b>strateg</b>	<b>P1</b>	<b>Source utilisée pour la correction</b>	<b>CHAMPS</b>	<b>strateg</b>	<b>P1</b>	<b>Source utilisée pour la correction</b>
Agriculture	A10	1,61	Reflet (Cereq)	Universités	U11	1,58	Couplage SISE
	A20	1,77			U12	1,43	
	A30	1,61			U13	1,40	
	A40	1,62			U14	1,54	
	A50	1,55			U15	1,41	
CFA	C1G	1,16	Enquête 51 (MEN)		U16	1,37	
	C1I	1,80			U17	1,13	
	C1T	1,41			U18	1,22	
	C2G	1,91			U21	1,06	
	C2I	1,70			U22	1,08	
	C2T	1,54			U23	1,21	
	C3I	1,45			U24	1,08	
	C3T	1,48			U25	1,14	
	C4I	1,40			U26	1,14	
	C4T	1,46			U27	1,10	
	C5I	1,42			U28	1,39	
	C5T	1,52			U31	1,04	
	C6I	1,49			U32	1,05	
C6T	1,53	U33	1,09				
Apprentis Univ.	G00	1,79	Couplage SISE	U34	1,15		
IUFM	I10	2,12	Source : MEN	U35	1,21		
	I20	1,38		U36	1,18		
Ministère Sports	J01 à J04	1,00	Source : MJSVA	U37	1,08		
Rectorat	R01 à R99	1,00	//	U43	1,15		
Ecoles du sup.	Z01	1,70	Source : MEN	U44	1,12		
	Z02	1,11	Source : MEN	U45	1,30		
	Z03	1,00	//	U46	1,14		
	Z04	1,00	//	U47	1,08		
	Z05	1,00	//	U53	1,10		
	Z06	1,00	//	U54	1,17		
	Z07	1,00	//	U55	1,18		
	Z08	1,00	//	U56	1,12		
	Z09	1,00	//	U57	1,09		
	Z10	1,20	Source : MEN	U58	1,36		

	Z11	1,00	//			U60	1,00	
	Z12	1,00	//			U63	1,24	
	Z13	1,00	//			U64	1,36	
DREES	s10	2,03	Source : FINESS (DREES)			U65	1,52	
	s20	2,93				U66	1,33	
	s30	3,42				U67	1,33	
	s40	1,34				U68	1,27	
	s50	1,70				U69	1,61	

***Coefficient P2 moyen par strate (variable Strateg)***

<b>strateg</b>	<b>P2</b>	<b>strateg</b>	<b>P2</b>	<b>strateg</b>	<b>P2</b>
C1G	4,89	U11	4,87	s10	1,21
C1I	4,12	U12	3,17	s20	1,11
C1T	3,82	U13	5,30	s30	1,10
C2G	4,30	U14	4,26	s40	1,00
C2I	2,79	U15	5,13	s50	1,00
C2T	3,73	U16	4,89	A10	3,85
C3I	3,41	U17	5,46	A20	3,82
C3T	3,40	U18	3,30	A30	4,55
C4I	3,56	U21	5,20	A40	3,82
C4T	3,20	U22	4,11	A50	3,27
C5I	2,39	U23	4,78	Z01	1,00
C5T	2,21	U24	5,28	Z02	4,41
C6I	2,42	U25	7,26	Z03	7,12
C6T	3,08	U26	5,63	Z04	6,96
R01	3,28	U27	5,96	Z05	7,72
R02	5,47	U28	3,86	Z06	6,64
R03	3,89	U31	1,27	Z07	7,80
R05	3,09	U32	1,25	Z08	7,80
R11	2,82	U33	1,24	Z09	6,85
R12	2,39	U34	1,10	Z10	5,33
R13	2,37	U35	1,10	Z11	7,80
R14	3,38	U36	1,15	Z12	7,46
R15	3,27	U37	1,29	Z13	6,64
R16	3,36	U43	5,61	J01	1,20
R17	3,82	U44	5,21	J02	1,00
R18	2,75	U45	5,62	J03	1,00
R19	2,62	U46	6,09	J04	1,00
R20	4,67	U47	6,70	G00	1,34
R21	6,43	U53	1,72	I10	5,08
R22	3,93	U54	1,62	I20	5,99
R23	4,02	U55	1,61		
R24	4,61	U56	1,80		
R25	5,39	U57	1,87		
R26	4,17	U58	1,13		
R27	3,00	U60	1,00		

R28	4,70		U63	1,00
R29	4,11		U64	1,00
R30	4,14		U65	1,00
R31	4,90		U66	1,00
R32	3,90		U67	1,00
R33	4,33		U68	3,83
R34	6,09		U69	5,46
R99	3,46			

## Modalités de la variable STRATEG

A10 = A10-163 - 164 BAC PROFESSIONNEL	R17 = R17-BEP Term-Tertia
A20 = A20-165 BTA (LYCÉE AGRICOLE)	R18 = R18-BEP Non Term-Indust
A30 = A30-166 - 170 BAC TECHNO	R19 = R19-BEP Non Term-Tertia
A40 = A40-167 - 168 CAPA BPA BEPA	R20 = R20-2de /1e S L ES
A50 = A50-162 BTSA (LYCÉE AGRICOLE)	R21 = R21-Term S L ES
C1G = C1G-Apprenti niveau 1 et 2 spe générale	R22 = R22-1e STI STL BT AdapT-Ind
C1I = C1I-Apprenti niveau 1 et 2 spe indus	R23 = R23-1e STT SMS BT AdapT-Tert
C1T = C1T-Apprenti niveau 1 et 2 spe tertiaire	R24 = R24-Term STI STL BT AdapT-Ind
C2G = C2G-Apprenti niveau 3 spe générale	R25 = R25-Term STT SMS BT AdapT-Tert
C2I = C2I-Apprenti niveau 3 spe indus	R26 = R26-Bac pro Non Term-Indust
C2T = C2T-Apprenti niveau 3 spe tertiaire	R27 = R27-Bac pro Non Term-Tertia
C3G = C3G-Apprenti niveau bac pro spe générale	R28 = R28-Bac pro Term-Indust
C3I = C3I-Apprenti niveau bac pro spe indus	R29 = R29-Bac pro Term-Tertia
C3T = C3T-Apprenti niveau bac pro spe tertiaire	R30 = R30-BTS Term-Indust
C4G = C4G-Apprenti autre niveau 4 spe générale	R31 = R31-BTS Term-Tertia
C4I = C4I-Apprenti autre niveau 4 spe indus	R32 = R32-BTS Non Term-Indust
C4T = C4T-Apprenti autre niveau 4 spe tertiaire	R33 =R33-BTS Non Term-Tertia
C5G = C5G-Apprenti CAP spe générale	R34 =R34-DECF DECF...
C5I = C5I-Apprenti CAP spe indus	R99 =R99-Inconnu
C5T = C5T-Apprenti CAP spe tertiaire	U11 =U11-L1-IUT Indust
C6G = C6G-Apprenti BEP autre niv 5 spe générale	U12 =U12-L1-IUT Tertiaire
C6I = C6I-Apprenti BEP autre niv 5 spe indus	U13 =U13-L1-Sces Fond Appl
C6T = C6T-Apprenti BEP autre niv 5 spe tertiaire	U14 =U14-L1-Droit-Sces Po
G00 = G00-Apprenti du sup	U1 5 =U1 5-L1 -Eco-Gestion
I10 = I1 0-IUFM sortants 1 ere année	U16 =U16-L1-Lettres-Arts
I20 = I20-IUFM passage en 2ème IUFM	U17 =U17-L1-Sces Hum\$Soc
J01 = J01-Sortants Brevets Etat	U18 =U18-L1-Santé
J02 = J02-Sortants BPJEPS	U21 =U21 -L2-IUT Indust
J03 = J03-Sortants BAPAAT	U22 =U22-L2-IUT Tertiaire
J04 = J04-Sortants BEATEP	U23 =U23-L2-Sces Fond Appl
R00 = R00-Hors Champ (5e 4e)	U24 =U24-L2-Droit-Sces Po
R01 = R01 -4e Aménagée/techno	U25 =U25-L2-Eco-Gestion
R02 = R02-3e	U26 =U26-L2-Lettres-Arts
R03 = R03-3e techno/insertion/cpa..	U27 =U27-L2-Sces Hum\$Soc
R05 = R05-SEGPA (6e-3e)	U28 =U28-L2-Santé
R11 = R11-SEGPA (CAP)	U31 =U31 -L3-IUT Indust
R12 = R12-CAP Non Term-Indust	U32 =U32-L3-IUT Tertiaire
R13 = R13-CAP Non Term-Tertia	U33 =U33-L3-Sces Fond Appl
R14 = R14-CAP Term-Indust	U34 =U34-L3-Droit-Sces Po
R15 = R15-CAP Term-Tertia	U35 =U35-L3-Eco-Gestion
R16=R16-BEP Term-Indust	U36 =U36-L3-Lettres-Arts

U37 =U37-L3-Sces Hum§Soc  
U38 =U38-L3-Santé  
U41 =U41-M1-IUT Indust  
U42 =U42-M1-IUT Tertiaire  
U43 =U43-M1-Sces Fond Appl  
U44 =U44-M1-Droit-Sces Po  
U45 =U45-M1 -Eco-Gestion  
U46 =U46-M1 -Lettres-Arts  
U47 =U47-M1-Sces Hum§Soc  
U48 =U48-M1-Santé  
U51 =U51 -M2-IUT Indust  
U52 =U52-M2-IUT Tertiaire  
U53 =U53-M2-Sces Fond Appl  
U54 =U54-M2-Droit-Sces Po  
U55 =U55-M2-Eco-Gestion  
U56 =U56-M2-Lettres-Arts  
U57 =U57-M2-Sces Hum§Soc  
U58 =U58-M2-Santé  
U60 =U60-D-CIFRE  
U61 =U61 -D-IUT Indust  
U62 =U62-D-IUT Tertiaire  
U63 =U63-D-Sces Fond Appl  
U64 =U64-D-Droit-Sces Po

U65 =U65-D-Eco-Gestion  
U66 =U66-D-Lettres-Arts  
U67 =U67-D-Sces Hum§Soc  
U68 =U68-D-Santé  
U69 =U69-Ingénieurs Univ  
Z01= Z01-Diplômés écoles artistiques  
Z02= Z02-Ecoles de commerce  
Z03 = Z03-Diplômés autres écoles - service  
Z04 = Z04-Diplômés autres écoles - industriel  
Z05 = Z05-Diplômés écoles agricoles (vétérinaires)  
Z06 = Z06-Sortants privés universtaires niv 1  
Z07 = Z07-Sortants privés universtaires niv 2  
Z08 = Z08-Sortants privés universtaires niv 3  
Z09 = Z09-Sortants centre privé d'enseignement  
Z10 = Z10-Ecoles d'ingénieur  
Z11 = Z11-Ecoles d'administrations publiques + ENS  
Z12 = Z12-IEP  
Z13 = Z13-Ecoles DGA  
s10 = s10-Diplôme d'infirmier  
s20 = s20-Diplôme d'aide-soignant  
s30 = s30-Autres diplômes de santé  
s40 = s40-Diplômes d'assistant de service social et Educ Spe  
s50 = s50-Autres diplômes social + Kiné

## **ANNEXE 2 : BILAN DES TESTS DE QUESTIONNAIRE**

---

### *Conception de deux questionnaires*

Pour la première fois, le Céreq a constitué deux questionnaires différents pour une même enquête :

- un questionnaire allégé ne détaillant que le premier emploi et la situation à la date de l'enquête, avec des modules « parcours scolaire » et « caractéristiques individuelles » restreints,
- et un questionnaire complet reprenant la description habituelle de toutes les situations, le parcours scolaire dans une version approfondie, et des modules supplémentaires en vue des interrogations suivantes, sur l'expérience et les perspectives professionnelles.

Le questionnaire allégé était destiné aux extensions. D'une durée initialement estimée à 12 minutes, il s'est finalement établi à 14 minutes d'entretien environ.

Le questionnaire complet a duré 26 minutes environ.

Dans le questionnaire complet, un module de 4 minutes destiné à la Dares portait spécifiquement sur les formations en alternance.

La gestion de ces deux questionnaires s'est avérée très lourde lors de toutes les phases de l'enquête.

### *Tests papier et modifications des questionnaires*

- *Questionnaire complet*

Le test du questionnaire a été réalisé en interne au Céreq avec une équipe de quatre vacataires, pendant deux semaines et demi (du 25 septembre au 10 octobre 2006, de 17h à 20h30, avec un samedi en journée). Le questionnaire a été mis en forme pour une présentation sur papier. Le module non-emploi, n'ayant fait l'objet d'aucun changement par rapport à l'enquête Génération 2001, n'a pas été intégré dans ce test. Les enquêteurs ont travaillé en binôme la première semaine, pour s'aider dans la gestion complexe des filtres et pouvoir minuter l'interrogation. Un total de 46 questionnaires exploitables dans le champ a été obtenu.

- *Questionnaire allégé*

Le questionnaire allégé était un extrait du questionnaire complet. Néanmoins, il est apparu opportun de réaliser un test spécifique sur ce questionnaire puisqu'il devait être posé à environ 30 000 individus d'extensions. L'intégralité du questionnaire a été testée. L'objectif était en particulier de tester l'enchaînement des questions et d'évaluer le temps de la passation. Le test a été réalisé en interne au Céreq avec une équipe de quatre vacataires, du 11 au 13 octobre 2006, principalement de 17h à 20h30. Le questionnaire a été mis en forme pour une présentation sur papier. Un total de 39 questionnaires exploitables dans le champ a été obtenu.

- Principaux réajustements

Les tests papiers ont permis de réajuster les questionnaires avant la réunion du comité du label du CNIS, qui a eu lieu le 20 novembre 2006. Les modifications ont concerné les thématiques suivantes :

- « Jobs » de vacances (Q53) : imprécision des modalités (souvent, parfois, jamais)  
→ modification des modalités en incluant des intervalles de fréquence.
- Questions d'opinion exclusives (EA24, P01A, EA79) : difficulté à choisir  
→ EA79 : ajout de la modalité « Normalement payé » et suppression du début de la question « Par rapport au travail que vous faites, ... ».
- Motifs de refus d'un emploi (RA5A) : les individus parlaient beaucoup et souhaitaient pouvoir répondre plusieurs modalités. Le choix d'une raison principale était souvent difficile et se faisait avec l'enquêteur.  
→ possibilité de réponses multiples pour RA5A.
- Module apprentissage : c'est très souvent le CFA (sa « cellule du recrutement ») qui trouvait l'entreprise d'apprentissage, et les jeunes le disaient spontanément en AP2A, alors que l'item n'apparaissait qu'en AP2C, filtrée.  
→ modification de la question AP2C.
- Temps de travail : certains déclaraient un temps partiel mais précisaient qu'ils travaillaient 35 heures ou plus. Ils répondaient quelquefois en nombre d'heures par semaine.  
→ ajout du nombre d'heures entre parenthèses pour les questions qui précisaient la durée du temps partiel.
- Salaires (EA61) : la diminution de salaire n'apparaissait pas dans les modalités de réponse, ce qui était problématique pour les personnes passées à temps partiel  
→ ajout d'un filtre orientant les personnes ayant changé de temps de travail (sur la période) vers les questions « montant du salaire ».  
→ suppression de la parenthèse « fixe + commissions » dans les questions sur le montant du salaire.
- Primes : les réponses n'étaient pas normalisées (les primes étaient quelquefois déclarées en euros par jour, en mois de salaire, annuellement, et très souvent en tranches car très variables). Les individus ne savaient pas toujours répondre (beaucoup de « ça dépend », « ça varie »). La question n'était pas toujours bien comprise (certains individus mentionnaient les primes de « fin de contrat »).  
→ concernant les primes, précision apportée : « hors avantages en nature ».
- Autres compléments de rémunération : ces questions étaient très rarement renseignées ; certains déclaraient des compléments en nature (tickets restaurant, voiture ...), d'autres incluaient les Assedic.  
→ suppression des questions relatives aux « autres compléments de rémunération »
- Activités d'encadrement (EA70) : « Est-ce que vous dirigez une ou plusieurs personnes ? »  
→ reformulation de la question comme suit : « Dans cet emploi, est-ce que vous dirigez (iez) des salariés ? ».

- Connaissances dans l'entreprise (EA13D) : le terme de relations extra-professionnelles n'était pas compris.  
→ reformulation de la modalité comme suit : « Par des amis ou de la famille ».
- Connaissance d'une embauche possible (EP15A) : les individus répondaient souvent « candidature spontanée » ou « démarches personnelles ». Un individu répond « pendant mon stage », un autre « suite à un boulot saisonnier » ou « suite à apprentissage ».  
→ ajout de la modalité « Candidature spontanée (NPC) ».  
→ modalité « par une relation personnelle » remplacée par « par une de vos relations » (personnelle ou professionnelle).
- Licenciement (EP81) : cas impossibles à coder : licenciement négocié, à l'amiable  
→ ajout d'une modalité « Autre ».
- Démission (EP83) : « Parce que j'avais trouvé un autre emploi » a été cité.  
→ reformulation de la question et des modalités.

## **ANNEXE 3 : OPTIMISER LE NOMBRE D'APPEL**<sup>33</sup>

---

Le Céreq souhaitait obtenir un échantillon d'environ 70 000 jeunes pour l'enquête « Génération 2004 ». Différents facteurs entrent en compte : la disponibilité de coordonnées téléphoniques, la capacité à joindre à la personne, l'acceptation de répondre de la personne jointe et son appartenance ou non au champ de l'enquête. Parmi les leviers, le nombre d'appel maximum réalisé pour tenter de joindre une personne, à des moments de la journée et de la semaine différents, peut jouer. Tant pour garantir le caractère pleinement aléatoire de l'échantillon que pour borner la durée de collecte et les coûts, il est en effet nécessaire de fixer un nombre maximum d'appels, que la maîtrise d'œuvre de la collecte doit réaliser avant de classer éventuellement l'individu comme injoignable.

Cette annexe présente un travail réalisé au cours de la collecte de l'enquête pour étudier le lien entre nombre d'appels et taux de réponse. Il cherche à prévoir le nombre d'individus qui devraient répondre au questionnaire en rehaussant à quinze le nombre maximum de tentatives d'appels. Il a aussi permis de voir comment la sortie d'un individu d'un état donné évolue au fil des appels. Pour cela, un modèle de durée non paramétrique à risques concurrents a été utilisé.

### ***Données nécessaires***

Nous avons choisi comme variable de temps le nombre d'appels effectués pour joindre chaque individu. La « survie » sera représentée par l'état « reste injoignable après le n-ième appel » et la « mort » par la sortie de cet état.

La « mort » elle-même sera représentée de plusieurs façons :

- « Enquête réalisée » (événement 1)
- « Hors champ ou hors cible » (événement 2)
- « Autres contacts utiles » (événement 3)

Tout individu restant injoignable mais qui est à retenter est donc considéré comme une donnée censurée.

Ces différents types de sorties sont dits « risques concurrents ».

Pour chaque version de l'enquête, allégée et complète, on distingue trois types de terrain :

Terrain1 : Les individus directement enquêtés et ceux issus des enrichissements classiques, en place lors des enquêtes antérieures.

Terrain 2 : Les individus injectés en enquête après une recherche nom+département.

---

<sup>33</sup> Document méthodologique réalisé en cours de collecte de l'enquête « Génération 2004 »

Terrain 3 : Les individus issus de la phase de préqualification.

Le modèle que nous utiliserons est donc un modèle non-paramétrique à risques concurrents.

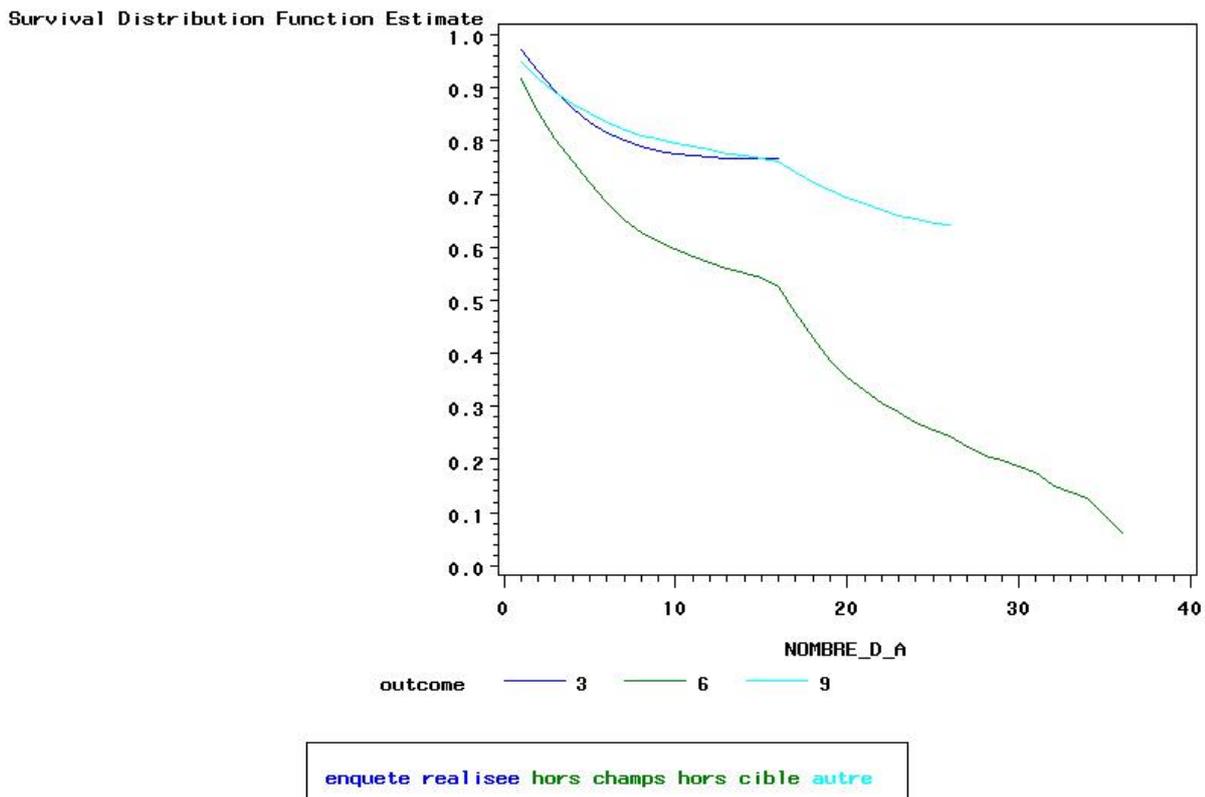
### *La survie des individus à chaque « risque »*

Les courbes ci-dessous représentent les courbes de survie aux deux versions de l'enquête réunies à chaque événement et ont été obtenues à l'aide d'une macroprocédure SAS mettant en œuvre l'estimation de Kaplan-Meier (procédure LIFETEST).

Prenons pour exemple la courbe de survie à l'évènement « hors champ ou hors cible » (vert), un individu qui survit à cet évènement ne survit pas forcément à tous les autres évènements. Contrairement aux modèles de durée sans risques concurrents, plusieurs raisons peuvent justifier la survie de l'individu à l'évènement 2 : soit il a connu l'évènement 1 . soit il a connu l'évènement 3 ; soit il est tout simplement resté injoignable. Il en est de même pour les deux autres évènements.

En observant ces courbes, on peut constater qu'un individu a plus de chances de survivre à l'évènement 1 ou 3 qu'à l'évènement 2 aux dix premiers appels. Mais ces courbes nous donnent uniquement les probabilités pour lesquelles un individu survit à tel ou tel évènement par opposition aux autres évènements, et ne donnent pas la probabilité pour laquelle un individu peut connaître cet évènement-là justement.

## Courbes de survie



### ***Probabilités pour que chaque individu connaisse l'un des évènements***

Pour connaître la probabilité que chaque individu puisse connaître l'un de ces évènements, il faut passer par les probabilités de survie suivantes :

- « Survivre à l'évènement 1 en prenant comme censures supplémentaires les évènements 2 et 3, en plus des données déjà censurées. » :  $P(1)$

- « Survivre à la fois aux évènements 1 et 2 en prenant comme censure supplémentaire l'évènement 3. » :  $P(1 \cup 2)$

- « Survivre aux trois évènements à la fois sans censure supplémentaire. » :  $P(1 \cup 2 \cup 3)$

Voici le tableau des valeurs de ces probabilités selon le nombre d'appels (jusqu'à 15) et le type d'évènement, les valeurs ayant été calculées à l'aide de la procédure LIFETEST du logiciel SAS :

<b>Nombre d'appels</b>	<b>Survie à 1</b>	<b>Survie à 1 et 2</b>	<b>Survie à 1, 2 et 3</b>
1	0,9724912146	0,8903495965	0,8386435228
2	0,9320191123	0,7924803321	0,7184335103
3	0,8936097005	0,7128427697	0,6262894771
4	0,8610974160	0,6476485757	0,5539274260
5	0,8341305534	0,5949913113	0,4969117850
6	0,8146033604	0,5493265381	0,4493886029
7	0,7997631029	0,5127463411	0,4120252610
8	0,7894018872	0,4878068664	0,3867026233
9	0,7814447529	0,4694000316	0,3683863658
10	0,7761249665	0,4549834602	0,3538016126
11	0,7723165927	0,4425222810	0,3412797320
12	0,7698705011	0,4321540951	0,3305488116
13	0,7678638051	0,4230711708	0,3208798828
14	0,7665994271	0,4158145414	0,3135854903
15	0,7658987319	0,4086777255	0,3060854260

Calculons maintenant les probabilités que l'individu soit classé dans chaque évènement au quinzième appel :  $P(1) = 0,7658987319$

Donc  $P(\bar{1}) = 1 - P(1) = 0,2341012681$ . C'est la probabilité pour que l'individu soit classé en « enquête réalisée ».

$P(1 \cup 2) = 0,4086777255$ , donc  $P(\bar{2}) = 1 - P(1 \cup 2) - P(\bar{1}) = 0,3572210064$  qui est la probabilité pour que l'individu soit classé à l'événement 2 donc en « hors champ ou hors cible ».

$P(1 \cup 2 \cup 3) = 0,3060854260$ . Cette probabilité correspond également à la probabilité pour que l'individu ne connaisse aucun des trois événements, c'est-à-dire qu'il leur survive et reste censuré.

Par ailleurs,  $P(\bar{3}) = [1 - P(1 \cup 2 \cup 3)] - [1 - P(1 \cup 2)] = 0,1025922995$  est la probabilité pour que l'individu soit classé à l'événement 3 donc en « autres contacts utiles ».

En sommant les trois probabilités avec celle qui correspond à la probabilité de rester censuré, on obtient 1.

### **Prévision sur les individus selon le type d'événement pour chaque terrain**

Avec la procédure FREQ du logiciel SAS et en utilisant les probabilités trouvées avec les individus censurés, on peut estimer pour chaque terrain d'enquête les effectifs, le nombre de personnes classées dans chaque événement, le nombre de personnes non classées qui pourraient connaître l'un des événements.

Dans les tableaux suivants, relatifs à chaque terrain d'enquête, les informations suivantes sont présentées :

*Effectif* : nombre total d'individus pour chaque événement

*eff1* : nombre d'individus susceptible de réaliser l'enquête + ceux qui l'ont déjà réalisé

*eff2* : nombre d'individus susceptibles d'être hors champ + ceux déjà classés en hors champ,

*eff3* : nombre d'individus susceptibles d'être en autres contacts utiles + ceux déjà classés en contacts utiles,

*eff4* : nombre d'individus susceptibles de rester censurés + ceux déjà censurés.

	<b>Terrain 1 : Individus directement enquêtés</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'évènement					
0	75208	17606	26866	7716	23020
1 (enquête réalisée)	27497	27497	0	0	0
2 (hors champ, hors cible)	26949	0	26949	0	0
3 (autres contacts utiles)	18503	0	0	18503	0
Total	148157	45103	53815	26219	23020

Pour le terrain « enquête directe », 148 157 individus ont été appelés au 15 mai 2007, soit deux mois après le début de l'enquête. Parmi eux, 75 208 n'ont pas décroché leur téléphone et sont donc à retenter. Au bout de quinze appels, 17 606 de ces individus devraient réaliser l'enquête, 26 866 devraient se retrouver en hors champ/hors cible, 7 716 devraient se retrouver en « autres contacts utiles » et 23 020 devraient rester injoignable. On devrait avoir 45 103 individus en « enquête réalisée », 53 815 en « hors champ », 26 219 en « autres contacts utiles ».

	<b>Terrain 2 : Individus enrichis après recherche nom+dép</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'évènement	9751	2283	3483	1000	2985
0					
1 (enquête réalisée)	4817	4817	0	0	0
2 (hors champ, hors cible)	15626	0	15626	0	0
3 (autres contacts utiles)	7505	0	0	7505	0
Total terrain 2	37699	7100	19109	8505	2985

Pour le terrain « recherche nom+département », il y a eu 37 699 individus appelés. Parmi eux, 9 751 sont à retenter. Au bout de quinze appels, 2 283 de ces individus devraient réaliser l'enquête, 3483 devraient se retrouver en hors champ/hors cible, 1 000 devraient se retrouver en « autres contacts utiles » et 2 985 devraient rester injoignable. Avec les individus déjà classés, on devrait se retrouver avec 7100 individus en « enquête réalisée », 19 109 en « hors champ », 8 505 en « autres contacts utiles ».

	<b>Terrain 3 : Individus pré qualifiés</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'évènement	48964	11463	17491	5023	14987
0					
1 (enquête réalisée)	5410	5410	0	0	0
2 (hors champ, hors cible)	37838	0	37838	0	0
3 (autres contacts utiles)	13041	0	0	13041	0
Total terrain 3	105253	16873	55329	18064	14987

Pour le terrain « pré qualification », il y a eu 105 253 individus appelés. Parmi eux, 48 964 sont à retenter. Au bout de quinze appels, 1 1463 de ces individus devraient réaliser l'enquête, 17 491 devraient se retrouver en hors champ, 5 023 devraient se retrouver en « autres contacts utiles » et 14 987 devraient rester injoignable. Avec les individus déjà classés, on devrait avoir 16872 individus en « enquête réalisée », 55 329 en « hors champ », 18 064 en « autres contacts utiles ».

	<b>Ensemble des trois terrains</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Terrain 1	148157	45103	53815	26219	23020
Terrain 2	37699	7100	19109	8505	2985
Terrain 3	105253	16873	55329	18064	14987
Total des terrains	291109	69076	128253	52788	40992

Au total, 291 109 individus auront été appelés. Ainsi, au bout de quinze tentatives maximum d'appels par individu, on devrait réaliser 69 076 enquêtes tous terrains et versions de l'enquête confondus soit pratiquement la taille voulue de l'échantillon.

### *Prévisions sur la version allégée de Génération 2004 seule*

Les données utilisées sont les mêmes que pour les deux versions de l'enquête réunies. La démarche pour calculer la probabilité qu'un individu connaisse tel ou tel événement, est également reprise.

	<b>Terrain 1 : Individus directement enquêtés</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'événement	31038	7990	11252	3226	8570
0					
1 (enquête réalisée)	11915	11915	0	0	0
2 (hors champ, hors cible)	12771	0	12771	0	0
3 (autres contacts utiles)	8827	0	0	8827	0
Total terrain 1	64551	19905	24023	12053	8570

Pour le terrain « enquête directe », il y a eu 64 551 individus appelés au 15 mai 2007. Parmi eux, 31 038 sont à retenir. Au bout de quinze appels, 7 990 de ces individus devraient réaliser l'enquête. Avec les individus déjà classés, on devrait avoir 19 905 individus en « enquête réalisée » au total.

	<b>Terrain 2 : Individus enrichis après recherche nom + dép</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'événement	2788	718	1010	290	770
0					
1 (enquête réalisée)	2179	2179	0	0	0
2 (hors champ, hors cible)	6212	0	6212	0	0
3 (autres contacts utiles)	4057	0	0	4057	0
Total terrain 2	15236	2897	7222	4347	770

Pour le terrain « recherche nom+département », parmi les 15 236 individus appelés, 2 788 sont à retenter. Au bout de quinze appels, 718 de ces individus devraient réaliser l'enquête. Avec les individus déjà classés, on devrait avoir 2 897 individus en « enquête réalisée ».

	<b>Terrain 3 : Individus pré qualifiés</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'événement	16568	4265	6006	1722	4575
0					
1 (enquête réalisée)	2745	2745	0	0	0
2 (hors champ, hors cible)	13369	0	13369	0	0
3 (autres contacts utiles)	4511	0	0	4511	0
Total terrain 3	37193	7010	19375	6233	4575

Pour le terrain « pré qualification », sur 37 193 individus appelés, 16 568 sont à retenter. Au bout de quinze appels, 4 265 de ces individus devraient réaliser l'enquête. Avec les individus déjà classés, on devrait avoir 7 010 individus en « enquête réalisée ».

	<b>Ensemble des trois terrains</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Terrain 1	64551	19905	24023	12053	8570
Terrain 2	15236	2897	7222	4347	770
Terrain 3	37193	7010	19375	6233	4575
Total des terrains	116980	29812	50620	22633	13915

Au total, 116 980 individus auront été appelés. On devrait réaliser 29 812 enquêtes tous terrains confondus pour la version allégée au bout de quinze appels.

*Prévisions sur la version complète de Génération 2004 seule*

	<b>Terrain 1 : Individus directement enquêtés</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'événement	44170	9587	15352	4526	14705
0					
1 (enquête réalisée)	15582	15582	0	0	0
2 (hors champ, hors cible)	14178	0	14178	0	0
3 (autres contacts utiles)	9676	0	0	9676	0
Total terrain 1	83606	25169	29530	14202	14705

Pour le terrain « enquête directe », il y a eu 44 170 individus appelés au 15 mai 2007. Au bout de quinze appels, 9 587 des individus censurés devraient réaliser l'enquête. Avec les individus déjà classés, on devrait avoir 25 169 individus en « enquête réalisée ».

	<b>Terrain 2 : Individus enrichis après recherche nom+dép</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'événement	6963	1511	2420	714	2318
0					
1 (enquête réalisée)	2638	2638	0	0	0
2 (hors champ, hors cible)	9414	0	9414	0	0
3 (autres contacts utiles)	3448	0	0	3448	0
Total terrain 2	22463	4149	11834	4162	2318

Pour le terrain « recherche élargie », 22 463 individus ont été appelés au 15 mai 2007. Au bout de quinze appels, 1 511 des individus censurés devraient réaliser l'enquête. Avec les individus déjà classés, on devrait avoir 4 149 individus en « enquête réalisée ».

	<b>Terrain 3 : Individus pré qualifiés</b>				
	<b>Effectif</b>	<b>eff1</b>	<b>eff2</b>	<b>eff3</b>	<b>eff4</b>
Type d'événement	32396	7032	11260	3320	10785
0					
1 (enquête réalisée)	2665	2665	0	0	0
2 (hors champ, hors cible)	24469	0	24469	0	0
3 (autres contacts utiles)	8530	0	0	8530	0
Total terrain 3	68060	9697	35729	11850	10785

Pour le terrain « pré qualification», il y a eu 68 060 individus appelés. Parmi eux, 32 396 sont à retenter. Au bout de quinze appels, 7 032 de ces individus devraient réaliser l'enquête, 11 260 devraient se retrouver en hors champ/hors cible, 3 320 devraient se retrouver en « autres contacts utiles » et 1 0785 devraient rester injoignables. Avec les individus déjà classés, on devrait avoir 9 697 individus en « enquête réalisée », 35 729 en « hors champ », 11 850 en « autres contacts utiles ».

	Ensemble des trois terrains				
	Effectif	eff1	eff2	eff3	eff4
Terrain 1	83606	25169	29530	14202	14705
Terrain 2	22463	4149	11834	4162	2318
Terrain 3	68060	9697	35729	11850	10785
Total des terrains	174129	39016	77092	30214	27807

Au total, 174129 individus auront été appelés. On devrait réaliser 39 016 enquêtes tous terrains confondus pour la version complète au bout de quinze appels pour le questionnaire complet.

En additionnant les effectifs d'enquête réalisée pour le questionnaire allégé et le questionnaire complet, on obtient  $29\ 812 + 39\ 016 = 68\ 828$  individus, soit 248 individus en moins par rapport à l'ensemble des deux questionnaires.

### *Conclusion*

En tentant quinze fois maximum chaque individu pour l'ensemble des questionnaires, on devrait arriver pratiquement à l'échantillon de répondants ciblé d'ici la date de fin de passation du questionnaire. On obtient à peu près le même résultat lorsqu'on applique le modèle de durée sur chaque version du questionnaire séparément.

La mise en œuvre du modèle a peut-être été effectuée trop tard dans la mesure où il était trop tard pour réajuster le volume d'individus en production à la hausse ou à la baisse. Il faudrait donc pour les prochaines enquêtes Génération utiliser ce modèle, de préférence, à la fin du premier mois d'enquête. De plus, il faudrait également faire une prévision sur les individus qualifiés avant le début de l'enquête afin de savoir combien d'entre eux passeront l'enquête.

D'autres paramètres comme l'origine du numéro de téléphone ou encore le jour de l'enquête pourraient être pris en compte par notre modèle ce qui n'est pas le cas ici en raison du manque de certaines données.



## **ANNEXE 4 : LES OUTILS DE LA SENSIBILISATION DES JEUNES**

---

*La lettre avis : cette lettre a été adressée avant la collecte aux jeunes dont l'adresse disponible était jugée suffisamment fiable.*

Madame, Mademoiselle, Monsieur,

Le Centre d'études et de recherches sur les qualifications (Céreq) mène une enquête nationale auprès des jeunes récemment sortis de formation. Cette étude vise à mieux connaître les parcours professionnels en début de vie active.

Pour cela, nous interrogeons, à partir du mois d'avril prochain, un échantillon de jeunes tiré au hasard. Vous faites partie de cet échantillon. Nous vous contacterons par téléphone dans les prochaines semaines. Avec votre accord, nous reconstituerons ensemble votre parcours depuis 2004.

Vos réponses serviront uniquement à produire des statistiques : elles resteront strictement anonymes et confidentielles.

Afin de faciliter notre futur entretien téléphonique, vous trouverez ci-joint un calendrier qui vous aidera à reconstituer votre itinéraire depuis 2004. Nous vous recommandons de le conserver afin de l'utiliser comme support lors de cet entretien.

Je vous remercie très vivement de votre participation à cette enquête.

Le Directeur du Céreq



Michel Quéré

- Si votre numéro de téléphone n'est plus le :

Merci de nous contacter au numéro vert suivant (appel gratuit depuis un téléphone fixe - en semaine entre 9h et 20h30, le samedi entre 10h et 18h) : **0 800 800 310**

en mentionnant la référence :

- Vous pouvez également utiliser ce numéro pour nous indiquer les moments les plus favorables pour vous joindre ou pour répondre directement à l'enquête.

Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue d'intérêt général et de qualité statistique sans avoir de caractère obligatoire. Label n° 2007X712AU du Conseil National de l'Information Statistique, valable pour l'année 2007. En application de la loi n°51-711 du 7 juin 1951, les réponses à ce questionnaire sont protégées par le secret statistique et destinées à la production de statistiques par le Céreq. La loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites et présentées en enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Ce droit peut être exercé auprès du Céreq, 10 place de la Joliette 13 567 Marseille Cedex 02.





Afin que l'entretien se déroule de manière fluide et confortable, et pour assurer une bonne compréhension de vos réponses, l'utilisation d'un téléphone fixe est préférable.

#### COMMENT A-T-ON OBTENU VOS COORDONNEES ?

Suite à l'avis favorable de la CNIL (Commission Nationale de l'Informatique et des Libertés), vos coordonnées nous ont été communiquées par l'établissement de formation dans lequel vous étiez inscrit en 2003-2004 et ont été actualisées avec l'annuaire France Telecom.



#### LA CONFIDENTIALITE DES DONNEES



Vos réponses au questionnaire sont couvertes par le secret statistique et ne donneront lieu qu'à des exploitations anonymes. Lors de l'étude, les noms, adresses et numéros de téléphone sont systématiquement supprimés du fichier de l'enquête.

#### L'OBJECTIF DE L'ENQUETE

Cette enquête cherche à décrire les situations professionnelles des jeunes quel que soit le diplôme préparé ou la formation suivie en 2003-2004.

Les résultats seront destinés à améliorer les choix d'orientation et à mieux adapter les formations aux réalités de l'emploi.



#### QU'EST-CE QUE LE CEREQ ? QUEL EST SON ROLE ?

Céreq



Le Centre d'études et de recherches sur les qualifications est un établissement public sous tutelle du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche et du ministère de l'Emploi, de la Cohésion sociale et du Logement.

Pôle public d'expertise au service des acteurs de la formation et de l'emploi, il allie production de statistiques, recherches, études et accompagnement d'actions. Le Céreq élargit ses collaborations avec le monde de la recherche comme avec les professionnels et les administrations locales.

Pour plus d'informations, consulter le site du Céreq : [www.cereq.fr](http://www.cereq.fr)



## POUR REMPLIR LE CALENDRIER (VOIR AU DOS) ET AINSI FACILITER L'ENTRETIEN TELEPHONIQUE

### ☞ Nous vous suggérons :

- de situer les périodes d'emploi en vous aidant de vos CV, contrats de travail et bulletins de salaire
- de vous repérer dans le temps en faisant appel aux événements qui ont marqué votre vie

### ☞ Quelques repères chronologiques pour la période 2004 – 2007 :



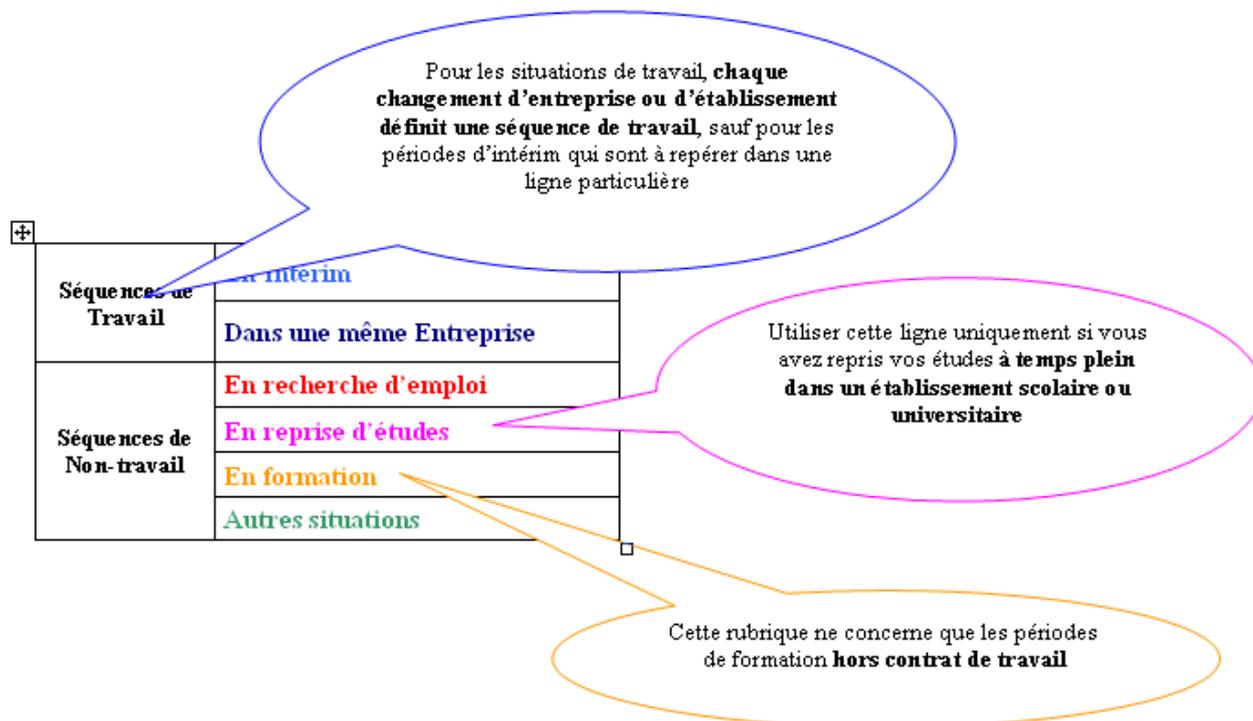
- Décembre 2004..... Tsunami en Asie du Sud
- Novembre 2005 ..... Emeutes dans les banlieues en France
- Juin 2006..... Coupe du Monde de football en Allemagne

### ☞ Quelques règles :

- Les diverses situations sont exclusives, c'est-à-dire que pour une période donnée, vous ne pouvez décrire qu'une seule situation.
- Si vous avez vécu plusieurs situations différentes pendant une même période, privilégiez toujours le travail sur les autres situations.
- Si, pendant un mois donné, vous avez travaillé trois semaines et le reste du mois vous étiez en recherche d'emploi, alors considérez le mois entier comme une période de travail.

### ☞ Quelques précisions sur les situations :

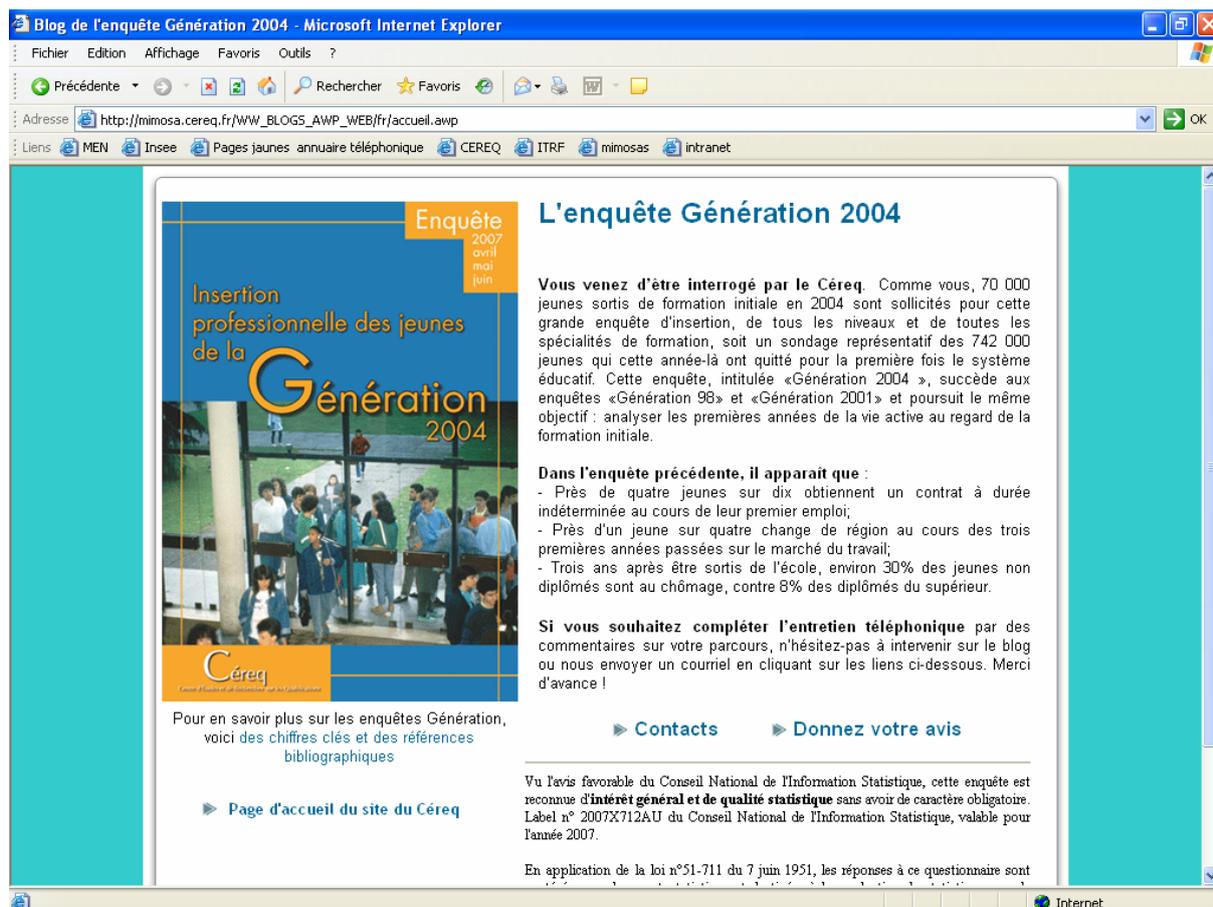
- On va découper votre parcours en différentes séquences, selon les situations que vous avez rencontrées :



## Le site internet et le blog « Génération 2004 »

Un « site » Internet a été mis en place au sein du site internet du Céreq, accessible dès la page d'accueil du site internet.

La page d'accueil de Génération 2004 présentait en quelques lignes les enquêtes Générations et décrivait plus précisément l'interrogation en cours :



**Blog de l'enquête Génération 2004 - Microsoft Internet Explorer**

Fichier Edition Affichage Favoris Outils ?

Précédente Recherche Favoris

Adresse [http://mimosas.cereq.fr/WW\\_BLOG5\\_AWP\\_WEB/fr/accueil.awp](http://mimosas.cereq.fr/WW_BLOG5_AWP_WEB/fr/accueil.awp) OK

Liens MEN Insee Pages jaunes annuaire téléphonique CEREQ ITRF mimosas intranet

### Enquête 2007 avril mai juin

## Insertion professionnelle des jeunes de la Génération 2004

### L'enquête Génération 2004

**Vous venez d'être interrogé par le Céreq.** Comme vous, 70 000 jeunes sortis de formation initiale en 2004 sont sollicités pour cette grande enquête d'insertion, de tous les niveaux et de toutes les spécialités de formation, soit un sondage représentatif des 742 000 jeunes qui cette année-là ont quitté pour la première fois le système éducatif. Cette enquête, intitulée «Génération 2004 », succède aux enquêtes «Génération 98» et «Génération 2001» et poursuit le même objectif : analyser les premières années de la vie active au regard de la formation initiale.

**Dans l'enquête précédente, il apparaît que :**

- Près de quatre jeunes sur dix obtiennent un contrat à durée indéterminée au cours de leur premier emploi;
- Près d'un jeune sur quatre change de région au cours des trois premières années passées sur le marché du travail;
- Trois ans après être sortis de l'école, environ 30% des jeunes non diplômés sont au chômage, contre 8% des diplômés du supérieur.

**Si vous souhaitez compléter l'entretien téléphonique** par des commentaires sur votre parcours, n'hésitez pas à intervenir sur le blog ou nous envoyer un courriel en cliquant sur les liens ci-dessous. Merci d'avance !

► [Contacts](#) ► [Donnez votre avis](#)

Pour en savoir plus sur les enquêtes Génération, voici des chiffres clés et des références bibliographiques

► [Page d'accueil du site du Céreq](#)

Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue d'**intérêt général et de qualité statistique** sans avoir de caractère obligatoire. Label n° 2007X712AU du Conseil National de l'Information Statistique, valable pour l'année 2007.

En application de la loi n°51-711 du 7 juin 1951, les réponses à ce questionnaire sont...

Trois liens ont été intégrés dans cette page. Le premier lien renvoyait vers les principaux résultats de la précédente enquête (Génération 2001), Le second renvoyait vers une page permettant d'envoyer un mail au Céreq et le dernier vers un blog afin que les enquêtés puissent poster des messages et échanger leurs impressions au sujet de l'enquête.

Ci-dessous, la page présentant les principaux résultats de l'enquête précédente.

Page d'accueil < Vos commentaires sur l'enquête Génération 2004 < Chiffres clés et bibliographie

Qui sommes-nous ? Nouveautés Bases de données Portail documentaire

Page d'accueil  
Qui sommes-nous ?  
Enquête «Génération 2004»  
Vos commentaires  
Bibliographie

Les enquêtes Génération

Des chiffres clés  
Quelques résultats significatifs

Cliquer sur l'image pour agrandir sa taille

**Le premier emploi en quelques indicateurs**

	Principaux contrats d'embauche				Tous contrats	Salaires mensuels nets médians (euros/mois)
	CDD	CDD2	Intérim	Contrat à durée indéterminée		
Non qualifiés	21	25	22	18	20	979
CAP ou BEP ou diplôme, 2 <sup>e</sup> ou 3 <sup>e</sup>	26	26	31	10	13	1 000
CAP ou BEP	34	27	18	14	16	998
Technique	20	36	14	14	27	920
Intermédiaire	16	21	22	11	7	1 030
Bac non diplômé	23	27	38	9	18	1 000
Bac professionnel ou technologique	21	27	21	16	17	999
Technique	26	26	15	22	22	930
Intermédiaire	15	24	23	8	6	1 080
Bac B ou Bac2 non diplômé	22	27	22	13	20	1 000
Baccalauréat	12	17	21	6	10	1 120
De la 1 <sup>re</sup> à la 3 <sup>e</sup> année totale	43	47	7	1	8	1 370
DEUG	24	38	10	8	17	1 110
BTS ou DUT tertiaire	29	38	20	8	12	1 040
BTS ou DUT industriel	23	29	16	5	6	1 180
2 <sup>e</sup> année	21	30	7	7	18	1 190
IUT/UTL/UST/Geodis	21	21	7	6	19	1 150
Master sciences exactes	22	20	7	2	17	1 270
3 <sup>e</sup> année	26	21	4	3	10	1 440
1 <sup>er</sup> / 2 <sup>e</sup> / 3 <sup>e</sup> année, Faculté de sciences	14	13	7	5	13	1 500
Master sciences économiques	49	41	6	1	12	1 080
Écoles d'ingénieurs	26	17	4	2	2	1 430
Ingénieur	30	27	19	11	15	1 820

Chiffres issus de l'enquête sur les emplois et les carrières des jeunes diplômés (en milliers de personnes) sur 26 000 individus.

Chiffres en milliers de personnes

Elle a été réalisée par le service informatique du Céreq. Les pages ont été intégrées une semaine après le début du terrain sur le site Internet du Céreq. Le site a utilisé la « technologie » des blogs pour la mise en ligne des messages. Celui-ci a été amélioré au fur et à mesure, avec notamment l'intégration d'une charte pour le blog.

- *Statistiques de fréquentation*

Sur les trois mois de terrain, le site a été fréquenté environ 3 000 fois.

L'adresse mail [G2004@cereq.fr](mailto:G2004@cereq.fr) a été principalement sollicitée en début de terrain, surtout suite à l'envoi des lettres avis. Le mail a été principalement utilisé par les individus habitant à l'étranger pour fournir leurs coordonnées.

### Le numéro vert

Un numéro vert a été mis en place par la société prestataire pour « Génération 2004 ». Ce numéro a été commandé spécifiquement pour cette opération, et était domicilié chez le prestataire. Les appels ont été pris en charge par une équipe de télé-enquêteurs et de superviseurs spécialement préparés.

Ce numéro vert a été imprimé par le Céreq sur l'ensemble des courriers adressés aux individus sélectionnés pour la phase d'enquête. D'autre part, il a été communiqué par les télé-enquêteurs lors des contacts téléphoniques, aux parents refusant de nous communiquer les coordonnées de l'individu recherché, ainsi qu'aux proches d'individus difficiles à joindre.

La ligne a été ouverte, chez le prestataire, dès l'envoi du mailing (le 27 mars 2007), et a été fermée à la fin de la période de généralisation (le 13 juillet 2007).

Les plages d'ouverture du numéro vert ont été :

- Du lundi au vendredi : de 9 heures à 20 heures 30
- Le samedi : de 10 heures à 18 heures.

Le nombre d'enquêteurs mis en place afin de gérer les appels entrants a été adapté au regard du trafic d'appels (entre 2 et 8 personnes).

En dehors des heures d'ouvertures, un répondeur a été mis en place, laissant la possibilité aux appelants de déposer un message (géré le lendemain par l'équipe).

Au total, 9 865 appels ont été reçus, dont 9 083 ont pu être traités (92%). 201 messages répondeurs ont été reçus et traités.

## ANNEXE 5 : REGROUPEMENT DE LA NES EN 114 POSITIONS EN 67 POSTES

NES (pour génération 2004)	Libellé	Nomenclature INSEE
A00	Agriculture, sylviculture, pêche	A01, A02, A03
B00	Industries agricoles et alimentaires	B01 à B06
C11	Industrie de l'habillement et des fourrures	C11
C12	Industrie du cuir et de la chaussure	C12
C20	Edition, imprimerie, reproduction	C20
C31	Industrie pharmaceutique	C31
C32	Industrie de la parfumerie et entretien	C32
C41	Fabrication de meubles	C41
C42	Fabrication de bijoux, instruments de musique	C42
C43	Fabrication d'articles de sport, de jeux, industries diverses	C43
C44	Fabrication d'appareils domestiques	C44
C45	Fabrication d'appareils de réception	C45
C46	Fabrication de matériel optique, photo, horlogerie	C46
D00	Industrie automobile	D01, D02
E10	Construction navale, aéronautique, ferroviaire, moto	E11 à E14
E20	Chaudronnerie, métallurgie pour la construction	E21, E22
E29	Fabrication de machines et équipements - dont armes	E23 à E28
E31	Fabrication de machines de bureau et de matériel informatique	E31
E32	Fabrication moteurs, génératrices, transformateurs électriques	E32
E33	Fabrication d'appareils d'émissions et de transmission	E33
E39	Fabrication de matériel médico-chirurgical et d'orthopédie, matériel de mesure-contrôle	E34, E35
F10	Industries des produits minéraux	F11, F12, F13
F20	Industrie textile	F21, F22, F23
F31	Industries du bois	F31
F39	Fabrication papier, carton, articles dérivés	F32, F33
F45	Caoutchouc	F45
F46	Plasturgie	F46
F49	Chimie, para-chimie, fibres artificielles et synthétiques	F41, F42, F43
F50	Métallurgie et transformation des métaux - dont récupération	F51 à F56
F61	Fabrication de matériel électrique	F61
F62	Fabrication de composants électroniques	F62
G99	Production de combustibles et de carburants - Eau, gaz, électricité	G11 à G15, G21, G22
H00	Construction	H01, H02
J10	Commerce et réparation automobile	J10
J20	Commerce de gros, intermédiaires	J20

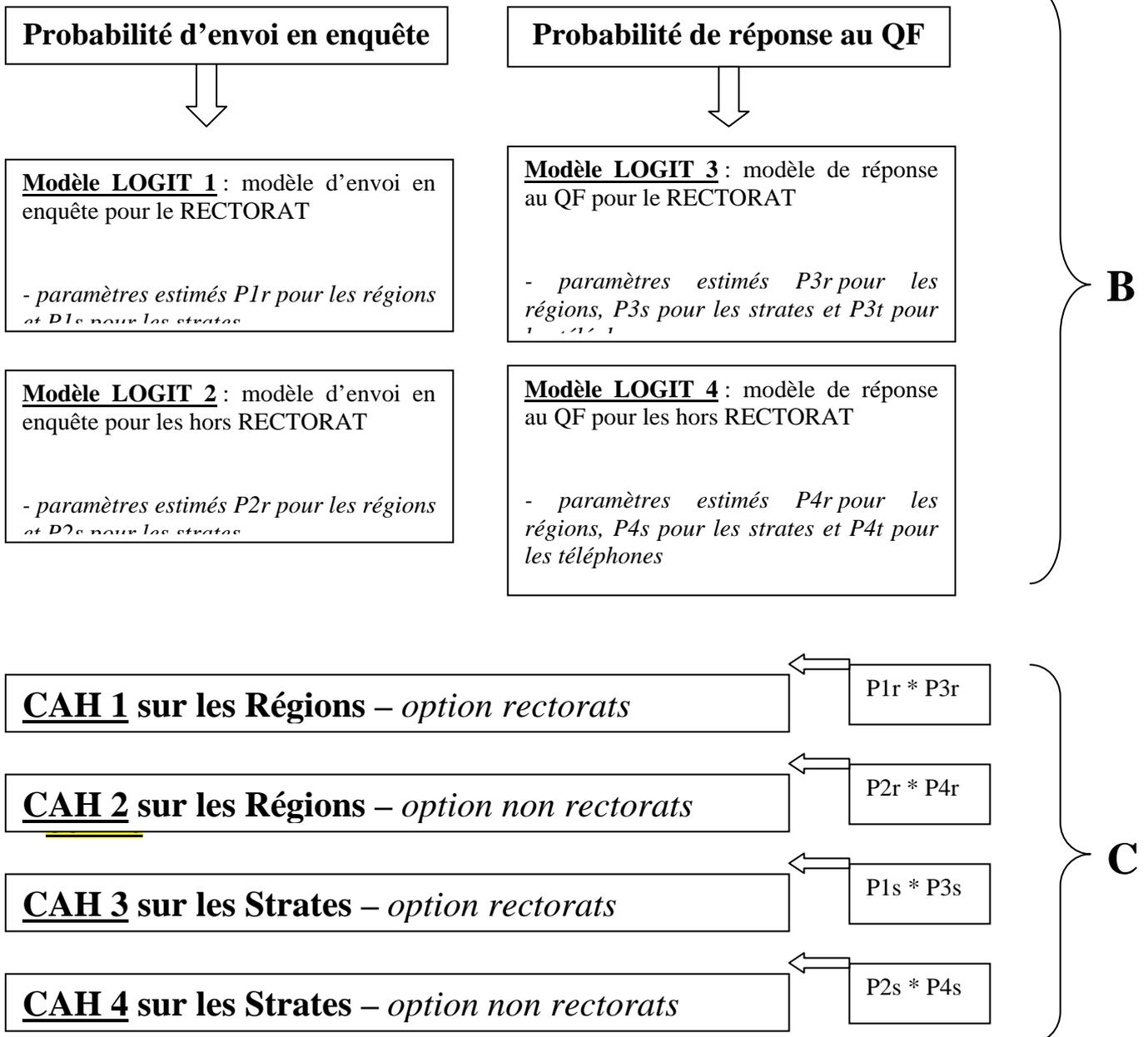
J31	Grande distribution à dominante alimentaire	J31
J32	Autres magasins alimentation	J32
J33	Autres commerces de détail, réparations	J33
K03	Transports routiers de marchandises	K03
K10	Autres transports, auxiliaires des transports	K01, K02, K04, K07, K08, K09
K05	Transports aériens, espace	K05, K06
L01	Banques, caisses d'épargne, société de crédit	L01
L02	Assurances	L02
L03	Auxiliaires financiers et d'assurances	L03
M09	Activités immobilières	M01, M02
N10	Postes et télécommunications	N11, N12
N21	Conseil et assistance informatique	N21
N22	Services juridiques, comptables, conseil en gestion, communication	N22
N23	Administration d'entreprise	N23
N24	Publicité, études de marché, sondages	N24
N25	Architecture, ingénierie, contrôle	N25
N30	Nettoyage	N33 74.7Z
N31	Location sans opérateur	N31
N32	Sélection et fourniture de personnel	N32
N33	Services divers fournis principalement aux entreprises	N33 74.8A à N33 74.8K
N34	Assainissement, voirie, gestion des déchets	N34
N35	Sécurité, gardiennage	N33 74.6Z
N40	Recherche et développement	N40
P10	Hôtels et restaurants	P10
P20	Activités récréatives, culturelles et sportives	P21, P22
P31	Services personnels	P31
P32	Services domestiques	P32
Q10	Education	Q10
Q21	Santé	Q21
Q22	Action sociale	Q22
R10	Administration publique	R10
R20	Activités associatives et extra-territoriales	R21, R22
Z99	Impossible à coder	

## ANNEXE 6 : LE REDRESSEMENT DE LA NON REPONSE TOTALE

Cette annexe complète la partie 5.3 relative à la pondération finale. Elle présente les résultats des modèles logistiques utilisés.

### SCHEMA GENERAL DE LA DEMARCHE

#### MODELES LOGIT => 3 variables discriminantes



## MODELES LOGISTIQUES UTILISES

### *LOGIT 1 : ENVOI EN ENQUETE – RECTORAT*

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
tel	2	8657.664 6	<.0001
region	21	977.7311	<.0001
strateg	28	1304.171 5	<.0001

Analyse des estimations de la vraisemblance maximum						
Paramètre		DF	Estimation	Erreur std	Khi 2 de Wald	Pr > Khi 2
Intercept		1	-0.6002	0.0128	2198.081 0	<.0001
tel	Prequal	1	-0.6120	0.0158	1493.599 8	<.0001
tel	Rech ad	1	1.1409	0.0123	8574.599 0	<.0001
region	21	1	0.0800	0.0428	3.4848	0.0619
region	22	1	-0.1559	0.0401	15.1331	0.0001
region	23	1	-0.0541	0.0400	1.8269	0.1765
region	24	1	-0.0636	0.0389	2.6709	0.1022
region	25	1	0.3083	0.0328	88.3784	<.0001
region	26	1	0.0621	0.0418	2.2127	0.1369
region	31	1	-0.1605	0.0265	36.6771	<.0001
region	41	1	0.0858	0.0286	9.0346	0.0026
region	42	1	0.0412	0.0423	0.9517	0.3293
region	43	1	0.2421	0.0437	30.6639	<.0001
region	52	1	0.0936	0.0307	9.3112	0.0023
region	53	1	0.0196	0.0338	0.3365	0.5619
region	54	1	0.2146	0.0315	46.3797	<.0001
region	72	1	-0.0232	0.0384	0.3645	0.5460
region	73	1	-0.00282	0.0393	0.0052	0.9428
region	74	1	0.3030	0.0500	36.6474	<.0001
region	82	1	-0.0615	0.0254	5.8539	0.0155

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>region</b>	<b>83</b>	1	0.1427	0.0448	10.1228	0.0015
<b>region</b>	<b>91</b>	1	-0.2198	0.0387	32.2028	<.0001
<b>region</b>	<b>93</b>	1	-0.2197	0.0279	61.8264	<.0001
<b>region</b>	<b>94</b>	1	-0.0872	0.0701	1.5477	0.2135
<b>strateg</b>	<b>R01</b>	1	0.2938	0.1196	6.0370	0.0140
<b>strateg</b>	<b>R02</b>	1	0.7927	0.0288	755.5400	<.0001
<b>strateg</b>	<b>R03</b>	1	0.3816	0.0442	74.5257	<.0001
<b>strateg</b>	<b>R05</b>	1	-0.3178	0.0434	53.6672	<.0001
<b>strateg</b>	<b>R11</b>	1	-0.0206	0.1142	0.0327	0.8566
<b>strateg</b>	<b>R12</b>	1	-0.2705	0.0678	15.9031	<.0001
<b>strateg</b>	<b>R13</b>	1	-0.2876	0.0789	13.2876	0.0003
<b>strateg</b>	<b>R14</b>	1	-0.1241	0.0460	7.2733	0.0070
<b>strateg</b>	<b>R15</b>	1	-0.0505	0.0491	1.0576	0.3038
<b>strateg</b>	<b>R16</b>	1	-0.1021	0.0280	13.3099	0.0003
<b>strateg</b>	<b>R17</b>	1	-0.0788	0.0296	7.0933	0.0077
<b>strateg</b>	<b>R18</b>	1	-0.1768	0.0433	16.7190	<.0001
<b>strateg</b>	<b>R19</b>	1	-0.3678	0.0429	73.6416	<.0001
<b>strateg</b>	<b>R20</b>	1	0.6358	0.0384	274.0862	<.0001
<b>strateg</b>	<b>R21</b>	1	0.7370	0.0327	507.6456	<.0001
<b>strateg</b>	<b>R22</b>	1	-0.1970	0.0902	4.7645	0.0291
<b>strateg</b>	<b>R23</b>	1	-0.1856	0.0658	7.9429	0.0048
<b>strateg</b>	<b>R24</b>	1	0.0288	0.0459	0.3926	0.5309
<b>strateg</b>	<b>R25</b>	1	-0.0465	0.0292	2.5362	0.1113
<b>strateg</b>	<b>R26</b>	1	-0.0403	0.0667	0.3640	0.5463
<b>strateg</b>	<b>R27</b>	1	-0.3576	0.0547	42.7097	<.0001
<b>strateg</b>	<b>R28</b>	1	0.0560	0.0348	2.5860	0.1078
<b>strateg</b>	<b>R29</b>	1	-0.00452	0.0297	0.0233	0.8788
<b>strateg</b>	<b>R30</b>	1	0.1167	0.0374	9.7095	0.0018
<b>strateg</b>	<b>R31</b>	1	-0.0819	0.0287	8.1180	0.0044
<b>strateg</b>	<b>R32</b>	1	-0.0421	0.0692	0.3700	0.5430
<b>strateg</b>	<b>R33</b>	1	-0.1746	0.0537	10.5770	0.0011
<b>strateg</b>	<b>R34</b>	1	0.0587	0.0936	0.3927	0.5309

**LOGIT 2 : ENVOI ENQUETE – HORS RECTORAT**

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
tel	3	6901.600 1	<.0001
region	21	1220.614 4	<.0001
strateg	84	2824.199 8	<.0001

Analyse des estimations de la vraisemblance maximum						
Paramètre		DF	Estimation	Erreur std	Khi 2 de Wald	Pr > Khi 2
Intercept		1	-0.4554	0.0164	774.9968	<.0001
tel	Numéro	1	-0.0318	0.0102	9.6364	0.0019
tel	Prequal	1	-0.2794	0.0107	678.2090	<.0001
tel	Rech ad	1	0.6475	0.0078 8	6744.864 9	<.0001
region	21	1	-0.0607	0.0393	2.3806	0.1228
region	22	1	0.1257	0.0357	12.3576	0.0004
region	23	1	-0.0513	0.0337	2.3178	0.1279
region	24	1	-0.0542	0.0305	3.1618	0.0754
region	25	1	0.0467	0.0273	2.9151	0.0878
region	26	1	-0.0168	0.0339	0.2447	0.6209
region	31	1	-0.0288	0.0208	1.9220	0.1656
region	41	1	0.0549	0.0227	5.8600	0.0155
region	42	1	-0.00022	0.0281	0.0001	0.9938
region	43	1	0.1480	0.0324	20.9365	<.0001
region	52	1	0.0926	0.0234	15.7211	<.0001
region	53	1	-0.0247	0.0244	1.0262	0.3111
region	54	1	0.1461	0.0232	39.6366	<.0001
region	72	1	-0.0512	0.0240	4.5478	0.0330
region	73	1	-0.0382	0.0250	2.3295	0.1269
region	74	1	0.2514	0.0358	49.3636	<.0001
region	82	1	-0.0674	0.0174	15.0801	0.0001
region	83	1	0.0885	0.0321	7.5782	0.0059

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>region</b>	<b>91</b>	1	-0.1520	0.0275	30.5738	<.0001
<b>region</b>	<b>93</b>	1	0.0216	0.0173	1.5457	0.2138
<b>region</b>	<b>94</b>	1	0.00598	0.0654	0.0084	0.9271
<b>strateg</b>	<b>A10</b>	1	0.4904	0.0615	63.4956	<.0001
<b>strateg</b>	<b>A20</b>	1	0.3791	0.0729	27.0520	<.0001
<b>strateg</b>	<b>A30</b>	1	0.6053	0.0909	44.3482	<.0001
<b>strateg</b>	<b>A40</b>	1	0.3149	0.0402	61.4335	<.0001
<b>strateg</b>	<b>A50</b>	1	0.4621	0.0518	79.4455	<.0001
<b>strateg</b>	<b>C1G</b>	1	-0.2212	0.2477	0.7980	0.3717
<b>strateg</b>	<b>C1I</b>	1	0.0527	0.0788	0.4484	0.5031
<b>strateg</b>	<b>C1T</b>	1	-0.2135	0.0585	13.3010	0.0003
<b>strateg</b>	<b>C2G</b>	1	-0.0954	0.7086	0.0181	0.8929
<b>strateg</b>	<b>C2I</b>	1	0.1388	0.0534	6.7543	0.0094
<b>strateg</b>	<b>C2T</b>	1	-0.0523	0.0483	1.1747	0.2784
<b>strateg</b>	<b>C3I</b>	1	0.2033	0.0484	17.6682	<.0001
<b>strateg</b>	<b>C3T</b>	1	-0.1939	0.0705	7.5548	0.0060
<b>strateg</b>	<b>C4I</b>	1	0.1093	0.0492	4.9297	0.0264
<b>strateg</b>	<b>C4T</b>	1	-0.0846	0.0496	2.9045	0.0883
<b>strateg</b>	<b>C5I</b>	1	-0.1531	0.0253	36.5025	<.0001
<b>strateg</b>	<b>C5T</b>	1	-0.2050	0.0370	30.6545	<.0001
<b>strateg</b>	<b>C6I</b>	1	0.0729	0.0347	4.4078	0.0358
<b>strateg</b>	<b>C6T</b>	1	-0.1372	0.0575	5.7015	0.0170
<b>strateg</b>	<b>G00</b>	1	0.1687	0.0463	13.2965	0.0003
<b>strateg</b>	<b>I10</b>	1	-0.00754	0.0499	0.0228	0.8800
<b>strateg</b>	<b>I20</b>	1	0.2435	0.0573	18.0641	<.0001
<b>strateg</b>	<b>J01</b>	1	0.0677	0.0340	3.9733	0.0462
<b>strateg</b>	<b>J02</b>	1	-0.1074	0.1350	0.6332	0.4262
<b>strateg</b>	<b>J03</b>	1	0.0255	0.0844	0.0914	0.7624
<b>strateg</b>	<b>J04</b>	1	0.2208	0.0563	15.3727	<.0001
<b>strateg</b>	<b>U11</b>	1	0.2688	0.0997	7.2663	0.0070
<b>strateg</b>	<b>U12</b>	1	0.2673	0.0707	14.2997	0.0002
<b>strateg</b>	<b>U13</b>	1	0.2162	0.0477	20.5426	<.0001
<b>strateg</b>	<b>U14</b>	1	-0.0645	0.0531	1.4765	0.2243
<b>strateg</b>	<b>U15</b>	1	0.0420	0.0576	0.5311	0.4662
<b>strateg</b>	<b>U16</b>	1	-0.0377	0.0395	0.9111	0.3398

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
strateg	U17	1	0.0946	0.0426	4.9333	0.0263
strateg	U18	1	0.3331	0.0508	42.9737	<.0001
strateg	U21	1	0.4680	0.0625	56.0516	<.0001
strateg	U22	1	0.3002	0.0462	42.3101	<.0001
strateg	U23	1	0.2086	0.0611	11.6678	0.0006
strateg	U24	1	0.0119	0.1015	0.0138	0.9065
strateg	U25	1	0.1653	0.1007	2.6970	0.1005
strateg	U26	1	0.00859	0.0772	0.0124	0.9114
strateg	U27	1	0.2001	0.0870	5.2905	0.0214
strateg	U28	1	-0.0380	0.0998	0.1450	0.7034
strateg	U31	1	0.6666	0.0410	264.1587	<.0001
strateg	U32	1	0.5720	0.0372	236.3662	<.0001
strateg	U33	1	0.4566	0.0337	183.3429	<.0001
strateg	U34	1	0.1620	0.0431	14.1577	0.0002
strateg	U35	1	0.2354	0.0380	38.3189	<.0001
strateg	U36	1	0.2168	0.0332	42.6873	<.0001
strateg	U37	1	0.4877	0.0312	245.0799	<.0001
strateg	U43	1	0.2659	0.0579	21.0746	<.0001
strateg	U44	1	-0.1084	0.0602	3.2442	0.0717
strateg	U45	1	0.1108	0.0546	4.1248	0.0423
strateg	U46	1	-0.0628	0.0580	1.1714	0.2791
strateg	U47	1	0.1831	0.0580	9.9564	0.0016
strateg	U53	1	0.3523	0.0310	129.1654	<.0001
strateg	U54	1	0.1824	0.0360	25.7331	<.0001
strateg	U55	1	0.2359	0.0327	52.0068	<.0001
strateg	U56	1	0.2273	0.0619	13.5038	0.0002
strateg	U57	1	0.3346	0.0391	73.2965	<.0001
strateg	U58	1	0.1102	0.0595	3.4323	0.0639
strateg	U63	1	-0.4972	0.0412	145.8955	<.0001
strateg	U64	1	-1.1412	0.0920	153.7872	<.0001
strateg	U65	1	-0.7237	0.1370	27.9199	<.0001
strateg	U66	1	-0.9276	0.1001	85.8727	<.0001
strateg	U67	1	-0.9703	0.0758	163.8736	<.0001
strateg	U68	1	-0.4422	0.0598	54.6458	<.0001
strateg	U69	1	-0.0467	0.0738	0.3995	0.5274

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>strateg</b>	<b>Z01</b>	1	0.00283	0.0315	0.0080	0.9285
<b>strateg</b>	<b>Z02</b>	1	-0.4688	0.0456	105.6938	<.0001
<b>strateg</b>	<b>Z03</b>	1	-0.4726	0.1735	7.4217	0.0064
<b>strateg</b>	<b>Z04</b>	1	-0.4599	0.2917	2.4859	0.1149
<b>strateg</b>	<b>Z05</b>	1	-0.1956	0.2385	0.6724	0.4122
<b>strateg</b>	<b>Z06</b>	1	-0.3270	0.2583	1.6023	0.2056
<b>strateg</b>	<b>Z07</b>	1	-0.3836	0.1608	5.6893	0.0171
<b>strateg</b>	<b>Z08</b>	1	0.1372	0.1841	0.5552	0.4562
<b>strateg</b>	<b>Z09</b>	1	0.3093	0.1785	3.0033	0.0831
<b>strateg</b>	<b>Z10</b>	1	-0.1272	0.0468	7.3783	0.0066
<b>strateg</b>	<b>Z11</b>	1	-0.9693	0.5064	3.6631	0.0556
<b>strateg</b>	<b>Z12</b>	1	-0.4337	0.1984	4.7785	0.0288
<b>strateg</b>	<b>Z13</b>	1	-0.7577	0.4909	2.3821	0.1227
<b>strateg</b>	<b>s10</b>	1	0.000523	0.0305	0.0003	0.9863
<b>strateg</b>	<b>s20</b>	1	-0.0712	0.0377	3.5762	0.0586
<b>strateg</b>	<b>s30</b>	1	0.0443	0.0654	0.4589	0.4981
<b>strateg</b>	<b>s40</b>	1	-0.2295	0.0478	23.0737	<.0001

**LOGIT 3 : REPONSE AU QUESTIONNAIRE FILTRE - RECTORAT**

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
region	21	2202.3764	<.0001
strateg	28	3031.5135	<.0001

Analyse des estimations de la vraisemblance maximum						
Paramètre		DF	Estimation	Erreur std	Khi 2 de Wald	Pr > Khi 2
Intercept		1	0.2411	0.00798	912.6681	<.0001
region	21	1	-0.0813	0.0294	7.6515	0.0057
region	22	1	-0.0203	0.0276	0.5431	0.4612
region	23	1	-0.0330	0.0276	1.4322	0.2314
region	24	1	0.1758	0.0285	38.1409	<.0001
region	25	1	-0.1586	0.0224	50.3593	<.0001
region	26	1	0.0106	0.0294	0.1290	0.7195
region	31	1	-0.0285	0.0183	2.4208	0.1197
region	41	1	-0.0757	0.0199	14.4333	0.0001
region	42	1	0.0161	0.0299	0.2894	0.5906
region	43	1	-0.0246	0.0306	0.6443	0.4221
region	52	1	0.5134	0.0250	421.2994	<.0001
region	53	1	0.5258	0.0282	348.7123	<.0001
region	54	1	-0.2975	0.0208	204.9580	<.0001
region	72	1	0.2669	0.0288	86.0451	<.0001
region	73	1	0.1113	0.0284	15.3590	<.0001
region	74	1	-0.4871	0.0313	242.6376	<.0001
region	82	1	0.3667	0.0193	359.7031	<.0001
region	83	1	-0.0366	0.0312	1.3711	0.2416
region	91	1	-0.0743	0.0254	8.5547	0.0034

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>region</b>	<b>93</b>	1	0.0965	0.0193	25.1216	<.0001
<b>region</b>	<b>94</b>	1	-0.8571	0.0395	469.9978	<.0001
<b>strateg</b>	<b>R01</b>	1	-0.5233	0.0700	55.8738	<.0001
<b>strateg</b>	<b>R02</b>	1	-0.3348	0.0167	401.5624	<.0001
<b>strateg</b>	<b>R03</b>	1	-0.4707	0.0257	335.9970	<.0001
<b>strateg</b>	<b>R05</b>	1	-0.2918	0.0267	119.0412	<.0001
<b>strateg</b>	<b>R11</b>	1	-0.1316	0.0727	3.2824	0.0700
<b>strateg</b>	<b>R12</b>	1	-0.2237	0.0425	27.6471	<.0001
<b>strateg</b>	<b>R13</b>	1	-0.3356	0.0481	48.7350	<.0001
<b>strateg</b>	<b>R14</b>	1	-0.0584	0.0313	3.4794	0.0621
<b>strateg</b>	<b>R15</b>	1	-0.0719	0.0332	4.6934	0.0303
<b>strateg</b>	<b>R16</b>	1	0.1100	0.0199	30.5663	<.0001
<b>strateg</b>	<b>R17</b>	1	-0.0129	0.0201	0.4079	0.5230
<b>strateg</b>	<b>R18</b>	1	-0.00382	0.0293	0.0170	0.8963
<b>strateg</b>	<b>R19</b>	1	-0.2384	0.0267	79.7950	<.0001
<b>strateg</b>	<b>R20</b>	1	-0.4892	0.0227	462.6366	<.0001
<b>strateg</b>	<b>R21</b>	1	-0.1654	0.0202	67.2739	<.0001
<b>strateg</b>	<b>R22</b>	1	0.0741	0.0636	1.3573	0.2440
<b>strateg</b>	<b>R23</b>	1	-0.0283	0.0444	0.4056	0.5242
<b>strateg</b>	<b>R24</b>	1	0.3503	0.0355	97.1646	<.0001
<b>strateg</b>	<b>R25</b>	1	0.3201	0.0218	215.8282	<.0001
<b>strateg</b>	<b>R26</b>	1	0.2005	0.0490	16.7652	<.0001
<b>strateg</b>	<b>R27</b>	1	0.0250	0.0368	0.4609	0.4972
<b>strateg</b>	<b>R28</b>	1	0.3836	0.0269	203.9218	<.0001
<b>strateg</b>	<b>R29</b>	1	0.2357	0.0217	117.9004	<.0001
<b>strateg</b>	<b>R30</b>	1	0.4762	0.0299	253.9443	<.0001
<b>strateg</b>	<b>R31</b>	1	0.4045	0.0222	331.8650	<.0001
<b>strateg</b>	<b>R32</b>	1	0.1969	0.0508	15.0227	0.0001

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>strateg</b>	<b>R33</b>	1	0.0753	0.0380	3.9295	0.0474
<b>strateg</b>	<b>R34</b>	1	0.4636	0.0754	37.8150	<.0001

**LOGIT 4 : REPONSE AU QUESTIONNAIRE FILTRE – HORS RECTORAT**

<b>Analyse des effets Type 3</b>			
<b>Effet</b>	<b>DF</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>region</b>	21	927.4330	<.0001
<b>strateg</b>	84	5233.7823	<.0001

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>Intercept</b>		1	-0.4554	0.0164	774.9968	<.0001
<b>tel</b>	<b>Numéro</b>	1	-0.0318	0.0102	9.6364	0.0019
<b>tel</b>	<b>Prequal</b>	1	-0.2794	0.0107	678.2090	<.0001
<b>tel</b>	<b>Rech ad</b>	1	0.6475	0.00788	6744.8649	<.0001
<b>region</b>	<b>21</b>	1	-0.0607	0.0393	2.3806	0.1228
<b>region</b>	<b>22</b>	1	0.1257	0.0357	12.3576	0.0004
<b>region</b>	<b>23</b>	1	-0.0513	0.0337	2.3178	0.1279
<b>region</b>	<b>24</b>	1	-0.0542	0.0305	3.1618	0.0754
<b>region</b>	<b>25</b>	1	0.0467	0.0273	2.9151	0.0878
<b>region</b>	<b>26</b>	1	-0.0168	0.0339	0.2447	0.6209
<b>region</b>	<b>31</b>	1	-0.0288	0.0208	1.9220	0.1656
<b>region</b>	<b>41</b>	1	0.0549	0.0227	5.8600	0.0155
<b>region</b>	<b>42</b>	1	-0.00022	0.0281	0.0001	0.9938
<b>region</b>	<b>43</b>	1	0.1480	0.0324	20.9365	<.0001
<b>region</b>	<b>52</b>	1	0.0926	0.0234	15.7211	<.0001
<b>region</b>	<b>53</b>	1	-0.0247	0.0244	1.0262	0.3111
<b>region</b>	<b>54</b>	1	0.1461	0.0232	39.6366	<.0001
<b>region</b>	<b>72</b>	1	-0.0512	0.0240	4.5478	0.0330
<b>region</b>	<b>73</b>	1	-0.0382	0.0250	2.3295	0.1269
<b>region</b>	<b>74</b>	1	0.2514	0.0358	49.3636	<.0001
<b>region</b>	<b>82</b>	1	-0.0674	0.0174	15.0801	0.0001
<b>region</b>	<b>83</b>	1	0.0885	0.0321	7.5782	0.0059
<b>region</b>	<b>91</b>	1	-0.1520	0.0275	30.5738	<.0001
<b>region</b>	<b>93</b>	1	0.0216	0.0173	1.5457	0.2138

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>region</b>	<b>94</b>	1	0.00598	0.0654	0.0084	0.9271
<b>strateg</b>	<b>A10</b>	1	0.4904	0.0615	63.4956	<.0001
<b>strateg</b>	<b>A20</b>	1	0.3791	0.0729	27.0520	<.0001
<b>strateg</b>	<b>A30</b>	1	0.6053	0.0909	44.3482	<.0001
<b>strateg</b>	<b>A40</b>	1	0.3149	0.0402	61.4335	<.0001
<b>strateg</b>	<b>A50</b>	1	0.4621	0.0518	79.4455	<.0001
<b>strateg</b>	<b>C1G</b>	1	-0.2212	0.2477	0.7980	0.3717
<b>strateg</b>	<b>C1I</b>	1	0.0527	0.0788	0.4484	0.5031
<b>strateg</b>	<b>C1T</b>	1	-0.2135	0.0585	13.3010	0.0003
<b>strateg</b>	<b>C2G</b>	1	-0.0954	0.7086	0.0181	0.8929
<b>strateg</b>	<b>C2I</b>	1	0.1388	0.0534	6.7543	0.0094
<b>strateg</b>	<b>C2T</b>	1	-0.0523	0.0483	1.1747	0.2784
<b>strateg</b>	<b>C3I</b>	1	0.2033	0.0484	17.6682	<.0001
<b>strateg</b>	<b>C3T</b>	1	-0.1939	0.0705	7.5548	0.0060
<b>strateg</b>	<b>C4I</b>	1	0.1093	0.0492	4.9297	0.0264
<b>strateg</b>	<b>C4T</b>	1	-0.0846	0.0496	2.9045	0.0883
<b>strateg</b>	<b>C5I</b>	1	-0.1531	0.0253	36.5025	<.0001
<b>strateg</b>	<b>C5T</b>	1	-0.2050	0.0370	30.6545	<.0001
<b>strateg</b>	<b>C6I</b>	1	0.0729	0.0347	4.4078	0.0358
<b>strateg</b>	<b>C6T</b>	1	-0.1372	0.0575	5.7015	0.0170
<b>strateg</b>	<b>G00</b>	1	0.1687	0.0463	13.2965	0.0003
<b>strateg</b>	<b>I10</b>	1	-0.00754	0.0499	0.0228	0.8800
<b>strateg</b>	<b>I20</b>	1	0.2435	0.0573	18.0641	<.0001
<b>strateg</b>	<b>J01</b>	1	0.0677	0.0340	3.9733	0.0462
<b>strateg</b>	<b>J02</b>	1	-0.1074	0.1350	0.6332	0.4262
<b>strateg</b>	<b>J03</b>	1	0.0255	0.0844	0.0914	0.7624
<b>strateg</b>	<b>J04</b>	1	0.2208	0.0563	15.3727	<.0001
<b>strateg</b>	<b>U11</b>	1	0.2688	0.0997	7.2663	0.0070
<b>strateg</b>	<b>U12</b>	1	0.2673	0.0707	14.2997	0.0002
<b>strateg</b>	<b>U13</b>	1	0.2162	0.0477	20.5426	<.0001
<b>strateg</b>	<b>U14</b>	1	-0.0645	0.0531	1.4765	0.2243
<b>strateg</b>	<b>U15</b>	1	0.0420	0.0576	0.5311	0.4662
<b>strateg</b>	<b>U16</b>	1	-0.0377	0.0395	0.9111	0.3398
<b>strateg</b>	<b>U17</b>	1	0.0946	0.0426	4.9333	0.0263
<b>strateg</b>	<b>U18</b>	1	0.3331	0.0508	42.9737	<.0001

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
strateg	U21	1	0.4680	0.0625	56.0516	<.0001
strateg	U22	1	0.3002	0.0462	42.3101	<.0001
strateg	U23	1	0.2086	0.0611	11.6678	0.0006
strateg	U24	1	0.0119	0.1015	0.0138	0.9065
strateg	U25	1	0.1653	0.1007	2.6970	0.1005
strateg	U26	1	0.00859	0.0772	0.0124	0.9114
strateg	U27	1	0.2001	0.0870	5.2905	0.0214
strateg	U28	1	-0.0380	0.0998	0.1450	0.7034
strateg	U31	1	0.6666	0.0410	264.1587	<.0001
strateg	U32	1	0.5720	0.0372	236.3662	<.0001
strateg	U33	1	0.4566	0.0337	183.3429	<.0001
strateg	U34	1	0.1620	0.0431	14.1577	0.0002
strateg	U35	1	0.2354	0.0380	38.3189	<.0001
strateg	U36	1	0.2168	0.0332	42.6873	<.0001
strateg	U37	1	0.4877	0.0312	245.0799	<.0001
strateg	U43	1	0.2659	0.0579	21.0746	<.0001
strateg	U44	1	-0.1084	0.0602	3.2442	0.0717
strateg	U45	1	0.1108	0.0546	4.1248	0.0423
strateg	U46	1	-0.0628	0.0580	1.1714	0.2791
strateg	U47	1	0.1831	0.0580	9.9564	0.0016
strateg	U53	1	0.3523	0.0310	129.1654	<.0001
strateg	U54	1	0.1824	0.0360	25.7331	<.0001
strateg	U55	1	0.2359	0.0327	52.0068	<.0001
strateg	U56	1	0.2273	0.0619	13.5038	0.0002
strateg	U57	1	0.3346	0.0391	73.2965	<.0001
strateg	U58	1	0.1102	0.0595	3.4323	0.0639
strateg	U63	1	-0.4972	0.0412	145.8955	<.0001
strateg	U64	1	-1.1412	0.0920	153.7872	<.0001
strateg	U65	1	-0.7237	0.1370	27.9199	<.0001
strateg	U66	1	-0.9276	0.1001	85.8727	<.0001
strateg	U67	1	-0.9703	0.0758	163.8736	<.0001
strateg	U68	1	-0.4422	0.0598	54.6458	<.0001
strateg	U69	1	-0.0467	0.0738	0.3995	0.5274
strateg	Z01	1	0.00283	0.0315	0.0080	0.9285
strateg	Z02	1	-0.4688	0.0456	105.6938	<.0001

<b>Analyse des estimations de la vraisemblance maximum</b>						
<b>Paramètre</b>		<b>DF</b>	<b>Estimation</b>	<b>Erreur std</b>	<b>Khi 2 de Wald</b>	<b>Pr &gt; Khi 2</b>
<b>strateg</b>	<b>Z03</b>	1	-0.4726	0.1735	7.4217	0.0064
<b>strateg</b>	<b>Z04</b>	1	-0.4599	0.2917	2.4859	0.1149
<b>strateg</b>	<b>Z05</b>	1	-0.1956	0.2385	0.6724	0.4122
<b>strateg</b>	<b>Z06</b>	1	-0.3270	0.2583	1.6023	0.2056
<b>strateg</b>	<b>Z07</b>	1	-0.3836	0.1608	5.6893	0.0171
<b>strateg</b>	<b>Z08</b>	1	0.1372	0.1841	0.5552	0.4562
<b>strateg</b>	<b>Z09</b>	1	0.3093	0.1785	3.0033	0.0831
<b>strateg</b>	<b>Z10</b>	1	-0.1272	0.0468	7.3783	0.0066
<b>strateg</b>	<b>Z11</b>	1	-0.9693	0.5064	3.6631	0.0556
<b>strateg</b>	<b>Z12</b>	1	-0.4337	0.1984	4.7785	0.0288
<b>strateg</b>	<b>Z13</b>	1	-0.7577	0.4909	2.3821	0.1227
<b>strateg</b>	<b>s10</b>	1	0.000523	0.0305	0.0003	0.9863
<b>strateg</b>	<b>s20</b>	1	-0.0712	0.0377	3.5762	0.0586
<b>strateg</b>	<b>s30</b>	1	0.0443	0.0654	0.4589	0.4981
<b>strateg</b>	<b>s40</b>	1	-0.2295	0.0478	23.0737	<.0001

**ANNEXE 7 : UNE EXPÉRIMENTATION DE RECHERCHE DE  
COORDONNÉES TÉLÉPHONIQUES POUR AMÉLIORER LE TAUX  
DE RÉPONSE A L'ENQUÊTE**

---

**« Améliorer le taux de réponse dans les enquêtes téléphoniques : une expérimentation  
de recherche de coordonnées téléphoniques sur l'enquête Génération 2004 »**

*Papier présenté par Bérangère Duploux (Céreq, DEEVA)  
aux journées de méthodologies statistiques, organisées par l'Insee en mars 2009*

## Introduction

1. Les enquêtes par téléphone
  - 1.1. Historique du taux d'équipement
  - 1.2. Les atouts des enquêtes téléphoniques
  - 1.3. Difficultés et limites des enquêtes téléphoniques
  
2. L'enjeu des recherches de coordonnées téléphoniques
  - 2.1. Présentation de l'enquête Génération 2004 du Céreq
  - 2.2. La base de sondage de Génération 2004
    - 2.2.1. *Des coordonnées non exhaustives*
    - 2.2.2. *La mobilité géographique des individus*
  
3. Le protocole d'enrichissement des coordonnées téléphoniques
  - 3.1. L'enrichissement des coordonnées téléphoniques dans l'enquête « Génération 98 »
  - 3.2. Une expérience niçoise
  - 3.3. Le protocole de Génération 2004
  
4. Quels résultats pour quel taux de réponse ?
  - 4.1. Un gain net de répondants
  - 4.2. Des profils d'individus qui diffèrent
  - 4.3. Quel redressement ?
  - 4.4. Un coût important

## Conclusion

## Introduction

Accompagnant la progression du taux d'équipement des foyers en téléphones filaires, les enquêtes téléphoniques se sont largement développées dans des domaines très variés. Les organismes statistiques publics, comme l'INSERM, l'INED ou encore l'INSEE se sont eux-mêmes appropriés ce mode de recueil depuis de nombreuses années.

Même si elles comportent divers inconvénients en comparaison d'entretiens en « face à face », les enquêtes par téléphone et le recours au système CATI (Collecte Assistée par Téléphone et Informatique) présentent des atouts indiscutables. En particulier, le faible coût des entretiens par téléphone permet d'interroger des échantillons plus importants. Néanmoins, les évolutions récentes du paysage téléphonique en France introduisent de nouvelles difficultés. Baisse du taux d'équipement en téléphone filaire, augmentation du nombre de foyers équipés exclusivement de téléphones mobiles et dégroupage biaisent la représentativité des échantillons téléphoniques issus des annuaires. Enfin, le développement du télémarketing accompagne une baisse progressive mais marquée des taux de réponse aux enquêtes téléphoniques.

L'enquête « Génération 2004 » du Céreq mise en œuvre au printemps 2007 n'échappe pas à ces évolutions. Mais une baisse du taux de réponse, en partie liée à ces évolutions, a pu être anticipée. Une nouvelle méthode d'enrichissement des coordonnées téléphoniques, basée sur le modèle des « enquêtes de voisinage », a donc été réalisée en amont de l'enquête. Quel a été l'impact de ces nouvelles recherches sur le taux de réponse à l'enquête et sur l'échantillon des répondants ? Quels enseignements tirer de cette expérimentation ?

## 1. Les enquêtes par téléphone

### *1.1. Historique du taux d'équipement*

La part des ménages équipés d'un téléphone filaire a rapidement progressé en France entre le début des années 60 et la fin des années 90. Les ménages français rattrapent ainsi un retard d'équipement significatif, par rapport notamment aux Etats-Unis. Entre 1963 et 1997, la part des ménages équipés d'un téléphone filaire passe donc de 9,3 % d'après l'INSEE à environ 95,8 % selon Médiamétrie. Cette généralisation du téléphone fixe a permis le développement des enquêtes par téléphone, qui offrent alors la possibilité de tirer (dans l'annuaire France Télécom) un échantillon représentatif au sens probabiliste de l'ensemble des ménages, y compris des ménages non équipés. Des études ont, en effet, pu montrer que les 5% des ménages qui ne sont pas équipés en téléphones filaires ne présentaient pas des caractéristiques sociodémographiques différentes de celles des autres ménages<sup>34</sup>.

Mais ces données sont remises en question par plusieurs évolutions à l'œuvre depuis la fin des années 90. Tout d'abord, on observe une baisse du taux d'équipement des ménages français en téléphone filaire depuis la fin des années 90. Cette baisse explique l'apparition d'une nouvelle population, celle des détenteurs exclusifs de téléphones mobiles. Médiamétrie estime que la part des ménages

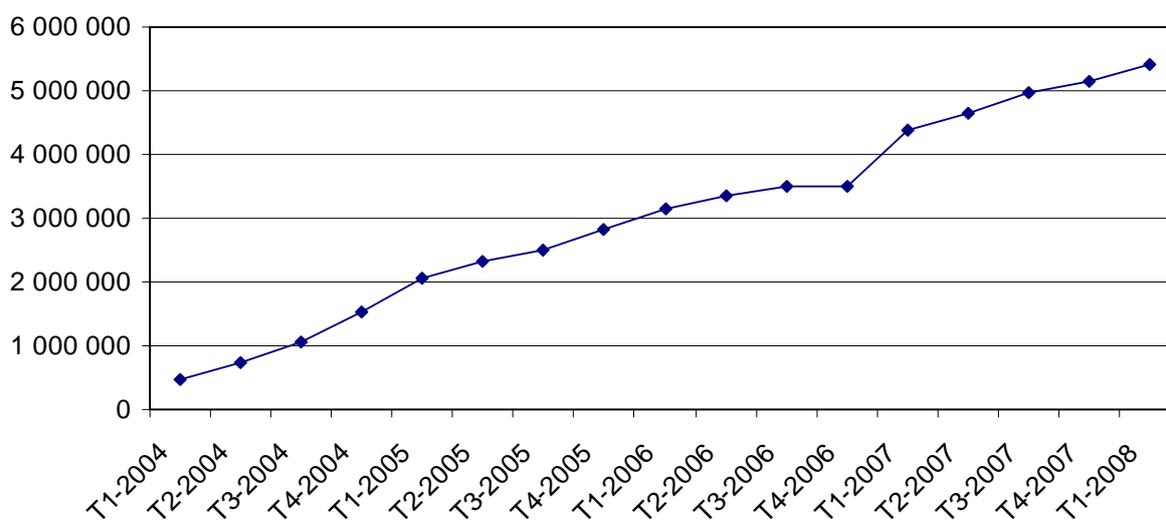
---

<sup>34</sup> Ambroise, P. et P. Mauris [1999], *L'usage du téléphone dans les sondages*

détenteurs exclusifs d'un téléphone mobile est passée de 6% en 2000 à 7 % en 2001 pour atteindre 9 % en 2002 et plus de 14 % en 2003<sup>35</sup>.

Enfin, si l'on semble noter une remontée du taux d'équipement des ménages en téléphone filaire ces toutes dernières années, il semble bien que ce soit sous l'effet du boom des lignes en dégroupage. D'après les données de l'ARCEP (Autorité de Régulation des Communications Electroniques et des Postes), le nombre de lignes dégroupées atteint 5,4 millions à la fin du premier trimestre 2008 et représente 15,6% de l'ensemble des lignes fixes.

**Graphique 1 : Evolution du nombre de lignes en dégroupage**



Source : ARCEP

### 1.2. Les atouts des enquêtes téléphoniques

L'enquête par téléphone présente un premier atout indiscutable qui est le coût réduit par rapport à une enquête en face-à-face et donc la possibilité de mettre en œuvre un sondage aléatoire simple sur des échantillons plus fournis : on obtient alors des indicateurs plus précis.

Les enquêtes téléphoniques sont aujourd'hui largement assistées par les systèmes CATI (Collecte Assistée par Téléphone et Informatique). La gestion et la passation des questionnaires (filtres et aiguillages divers) sont ainsi optimisées, les fichiers de numéros de téléphones sont automatiquement brassés grâce à des prédictifs d'appel. Enfin, tous les enquêteurs sont réunis sur un même plateau d'enquête, ce qui rend plus efficaces la formation et le contrôle.

<sup>35</sup> En 2004 : 14 %, en 2005 : 16 %, en 2006 : 15 % - Source Crédoc

Menée par l'INED en 1992, l'enquête ACSF a démontré, suite à différents tests, que l'entretien téléphonique présentait de nombreux avantages en comparaison du questionnement en face-à-face, y compris sur le sujet délicat des pratiques sexuelles. En fait, il semble que les individus de tous âges aient socialement adopté le téléphone, devenu en quelques années un vecteur banalisé de communication.

### ***1.3. Difficultés et limites des enquêtes téléphoniques***

Il existe 2 possibilités dans les enquêtes par téléphone :

- soit, l'annuaire téléphonique est la base de sondage elle-même
- soit, l'annuaire téléphonique doit permettre de retrouver les coordonnées téléphoniques d'individus présents dans une base de sondage à part et qui contient à minima leurs noms, prénoms et adresses (c'est le cas des enquêtes Générations du Céreq)

Dans les deux cas de figure, la question de l'exhaustivité du ou des annuaires se pose. Si cet annuaire ne contient qu'une partie des individus, l'échantillonnage sera d'autant plus biaisé que les « absents » de l'annuaire présentent des caractéristiques sociodémographiques marquées.

Le recours à l'annuaire mérite toutefois d'être justifié. On pourrait tout à fait envisager, dans le cas où l'annuaire est la base de sondage elle-même et ne sert donc pas à enrichir une autre base de sondage, de générer des nombres aléatoires pour constituer un échantillon de numéros de téléphone. Mais ce procédé se révèle rapidement coûteux sur de gros échantillons puisque de nombreux numéros ne sont pas attribués. De plus, comme l'expliquent Ambroise et Mauris « même si la structure des numéros à 10 chiffres reflète quelque peu la géographie (6 premières positions), il n'y a pas véritablement de bijection avec les communes, ce qui interdit de stratifier un échantillon ainsi tiré sur le critère commune »<sup>36</sup>.

Finalement, l'annuaire apparaît comme un outil indispensable et, jusqu'il y a peu, performant pour la mise en œuvre d'enquêtes téléphoniques. Jusqu'à la fin des années 90, l'annuaire France Télécom faisait référence : plus de 95 % des français étaient équipés de lignes fixes, lesquelles, conséquence du monopole de France Télécom, apparaissaient dans les pages blanches. La seule difficulté posée par cet annuaire était la présence d'une liste rouge composée d'individus équipés mais ne souhaitant pas apparaître dans l'annuaire. Cette difficulté a pu être contournée par la génération aléatoire de numéros (cf. travaux de Fréjean en 1990 et Tassi en 1992).

Plusieurs difficultés nouvelles sont toutefois apparues avec l'évolution du marché des télécommunications français.

Le développement de la téléphonie mobile, et donc l'apparition d'une population de détenteurs exclusifs de portables, a amputé d'autant l'annuaire France Télécom. Roy et Vanheuverzwyn en 2000 démontrent que cette population de détenteurs exclusifs de mobiles présentent des caractéristiques

---

<sup>36</sup> Ambroise, P. et P. Mauris [1999], *L'usage du téléphone dans les sondages*

particulières qui remettent en question la représentativité d'une enquête sur téléphone fixe. Dans l'enquête EROPP en 2002, l'Observatoire Français des Drogues et Toxicomanies (OFDT) a constitué un sous-échantillon de détenteurs exclusifs de mobiles (environ 200 individus). Cet échantillon présente un « profil sociodémographique très spécifique ». Il s'agit plus souvent d'hommes, plutôt jeunes et diplômés. Au final, l'opinion sur les drogues des enquêtés « portables » diffère significativement de celle des enquêtés « lignes fixes ».

Enfin, le développement des lignes dégroupées concurrentes à France Télécom entame encore l'exhaustivité des fameuses « pages blanches ». Finalement, il n'existe plus un seul annuaire France Télécom mais une multitude de sources, toutes incomplètes. La mise en place d'un annuaire universel, prévue par la loi n° 96-659 du 26 juillet 1996, devrait résoudre les difficultés liées au dégroupage et aux mobiles. Mais, douze ans après le vote de cette loi, ce service d'annuaire universel peine encore à convaincre : en cause, la faible participation de certains opérateurs à cet annuaire.

**Tableau 1 : Participation des opérateurs français de téléphonie à l'annuaire universel**

Septembre 2008	Nombre de numéros inscrits dans les listes (ordre de grandeur)	Taux de numéros inscrits dans les listes	Taux de numéros pour lesquels l'abonné a refusé la publication	État de la mise à disposition
<b>Service fixe</b>				
B3G	moins de 20.000	2% à 5%	<0,1%	
Colt <sup>(1)</sup>	moins de 20.000	50% à 100%	10% à 20%	
Complete! <sup>(2)</sup>				
France Télécom	plus de 20.000.000	50% à 100%	10% à 20%	
Free <sup>(1)</sup>	2.500.000 à 10.000.000	50% à 100%	10% à 20%	
Hub Telecom	moins de 20.000	1% à 2%	50% à 100%	
NeufCegetel <sup>(3)</sup>	500.000 à 2.500.000	50% à 100%	10% à 20%	
Noos	100.000 à 500.000	50% à 100%	20% à 50%	
Numericable*	100.000 à 500.000	50% à 100%	20% à 50%	
Outremer Telecom*	20.000 à 100.000	50% à 100%	10% à 20%	
FRNet 2 (ex Tele 2 France)*	500.000 à 2.500.000	50% à 100%	10% à 20%	
Telecom Italia	500.000 à 2.500.000	50% à 100%	20% à 50%	
Verizon	20.000 à 100.000	50% à 100%	20% à 50%	
<b>Service mobile</b>				
Amobile**	moins de 20.000	5% à 10%	50% à 100%	
Bouygues Telecom	100.000 à 500.000	1% à 2%	0,1% à 1%	
Carrefour Mobile**	moins de 20.000	<0,1%	50% à 100%	
Debitel*	moins de 20.000	1% à 2%	2% à 5%	
Digicel Antilles Françaises Guyane	moins de 20.000	<0,1%	<0,1%	
Neuf Cegetel	moins de 20.000	2% à 5%	<0,1%	
NRJ mobile*	moins de 20.000	<0,1%	<0,1%	
Breizh Mobile <sup>(4)</sup>	moins de 20.000	2% à 5%	50% à 100%	
Virgin Mobile <sup>(4)</sup>	20.000 à 100.000	5% à 10%	50% à 100%	
Orange Caraïbe	moins de 20.000	0,1% à 1%	20% à 50%	
Orange France	100.000 à 500.000	0,1% à 1%	<0,1%	
Orange Réunion	moins de 20.000	0,1% à 1%	1% à 2%	
Outremer Telecom*	moins de 20.000	1% à 2%	50% à 100%	
SFR	500.000 à 2.500.000	5% à 10%	20% à 50%	
SRR	20.000 à 100.000	5% à 10%	<0,1%	
Tele 2 mobile*	moins de 20.000	2% à 5%	<0,1%	

<sup>(1)</sup> Opérateurs ayant pris en compte des nombres de contrats d'abonnement et non des nombres de numéros

<sup>(2)</sup> L'opérateur s'est opposé à la publication de certaines informations

<sup>(3)</sup> les données de NeufCegetel intègrent les données relatives à AOL, Club-Internet

<sup>(4)</sup> Précédemment regroupés sous Omer Télécom

<sup>(\*)</sup> Opérateur intégré à l'édition du tableau de bord portant sur le mois de novembre 2006

<sup>(\*)</sup> Opérateur intégré à l'édition du tableau de bord portant sur le mois de novembre 2007

État de la mise à disposition

	Pas de disponibilité effective d'une offre de mise à disposition par l'opérateur de sa liste
	Disponibilité effective d'une offre mais aucune mise à disposition effectivement réalisée
	Mise à disposition de listes effectivement réalisée auprès d'un ou deux éditeurs
	Mise à disposition de listes effectivement réalisée auprès de plusieurs éditeurs
	Mise à disposition des listes effectivement réalisée très largement

Source : ARCEP

## 2. L'enjeu des recherches de coordonnées téléphoniques

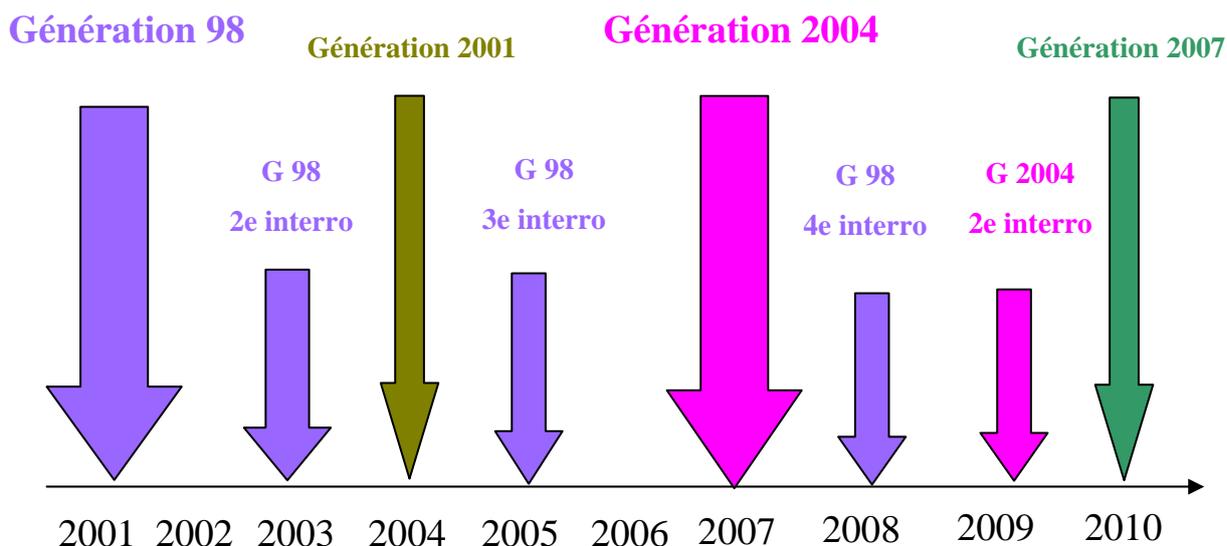
### 2.1. Présentation de l'enquête Génération 2004 du Céreq

Créé au début des années 1970 puis devenu établissement public en 1985, le Centre d'Etudes et de Recherches sur les Qualifications est placé sous la tutelle du Ministère de l'Education nationale, du Ministère de l'Economie, de l'Industrie et de l'Emploi et enfin du Ministère du Travail, des Relations sociales, de la Famille, de la Solidarité et de la Ville. Le Céreq a pour missions de développer des études et des recherches, collecter et exploiter des données originales dans le domaine de la relation formation-emploi, et formuler des avis et des propositions destinés à éclairer les choix en matière de politiques de formation.

L'enquête Génération 2004 s'inscrit dans le cadre du dispositif d'enquêtes de l'Observatoire National des Entrées dans la Vie Active (ONEVA). Les enquêtes « générationnelles » du Céreq s'adressent à une nouvelle génération de sortants du système éducatif tous les trois ans. On distingue deux types d'enquêtes « générationnelles » :

- les enquêtes générations « pleines » avec un échantillon minimum de 40 000 individus. Ces générations sont suivies en longitudinal et sont donc réinterrogées à intervalle régulier. On met ainsi en œuvre une première interrogation trois ans après la sortie du système éducatif, principalement tournée vers la production d'indicateurs et des usages institutionnels puis plusieurs ré-interrogations plutôt centrées sur les usages analytiques (notamment sur la question des parcours, des mobilités sur moyen terme), permettant en outre d'approfondir certains constats de la première étape (par exemple sur les insertions les plus problématiques) et d'analyser les trajectoires longues.
- les enquêtes « mini-générationnelles » avec un échantillon réduit d'environ 10 000 individus, interrogés une seule fois trois ans après la sortie du système éducatif et sur la base d'un questionnaire proche de celui des enquêtes générations « pleines ». Ces enquêtes ont vocation à réactualiser les indicateurs d'insertion sur les 3 premières années d'insertion

L'enchaînement des enquêtes Générations est donc le suivant :



L'enquête « Génération 2004 – interrogation à 3 ans » a eu lieu au printemps 2007 : elle interrogeait les sortants de formation initiale durant l'année scolaire 2003-2004 et concernait tous les niveaux et champs de la formation initiale.

Les différentes étapes de l'enquête Génération 2004 :

- Constitution de la base de sondage : la base de sondage de Génération 2004 a été constituée par collecte centralisée des fichiers d'élèves du rectorat et par une collecte établissement par établissement de tous les fichiers d'élèves des autres champs. L'objectif de cette collecte est de couvrir au mieux l'ensemble des formations initiales dispensées en France métropolitaine. On procède à un appariement des fichiers entre eux pour retirer de la base les individus repérés en poursuite d'étude.
- Tirage de l'échantillon : un tirage stratifié par régions et par niveau\*spécialité de formation a permis de constituer un échantillon de plus de 400 000 individus.
- **Enrichissement des coordonnées téléphoniques des individus de l'échantillon. Les informations contenues dans la base de sondage étant incomplètes, il convient de retrouver via des recherches de coordonnées téléphoniques un maximum de coordonnées et ce afin de minimiser la non-réponse à l'enquête.**
- Plateau d'enquête : jusqu'à 160 télé-enquêteurs ont mené des entretiens par téléphone du 2 avril au 12 juillet 2007. Le Céreq a ainsi collecté 65 300 questionnaires d'enquêtes.
- Apurements, codages et pondération : après réception des questionnaires au Céreq, les bases sont apurées et contrôlées. Des variables de profession, secteur d'activité et géo localisation sont créées. Enfin, il convient alors de calculer une pondération.
- Diffusion et exploitation des données : après plusieurs mois de contrôles et de travail, les bases peuvent enfin être diffusées. Les chargés d'études prennent en main la phase d'exploitation.

### Quelques publications liées à l'enquête « Génération 2004 » :

- « Génération 2004, des jeunes pénalisés par la conjoncture », Olivier Joseph, Alberto Lopez et Florence Ryk (Céreq). *Bref*, n° 248, janvier 2008.
- « Être diplômé de l'enseignement supérieur, un atout pour entrer dans la vie active ». Julien Calmand et Pierre Hallier (Céreq). *Bref*, n° 253, juin 2008
- « Quand l'école est finie » ou les premiers pas de la Génération 2004 dans la vie active. Ouvrage collectif.

## **2.2. La base de sondage de Génération 2004**

La base de sondage de Génération 2004 se heurte, comme toute enquête téléphonique, à deux grandes difficultés. La première difficulté consiste à collecter une base la plus exhaustive possible. Des efforts sont fournis dans ce sens par le Céreq en amont et pendant la collecte de la base de sondage. Pourtant, l'information collectée n'est pas toujours complète. Enfin, deuxième difficulté : entre le début de l'année scolaire 2004, c'est-à-dire en septembre 2003, lorsque les élèves transmettent les informations les concernant (et notamment leurs coordonnées téléphoniques) à leur établissement et le plateau d'enquête au printemps 2007, il s'est écoulé environ 3 ans et demi. La mobilité géographique des jeunes entame alors largement la fraîcheur de nos données.

### 2.2.1. Des coordonnées non exhaustives

A l'issue du processus de collecte de la base de sondage de Génération 2004, seuls 27 % des individus de la base de sondage possèdent un numéro de téléphone.

**Tableau 2 : Remplissage des variables de téléphone dans la base de sondage de Génération 2004**

Type d'établissement de sortie	Part des sortants du champ	Remplissage des variables de téléphone	
		aucun numéro	un numéro
Rectorats	53 %	aucun numéro : 53 %	un numéro : 0 %
		aucun numéro : 6 %	un numéro : 19 %
Centres de Formation des Apprentis	10 %	aucun numéro : 7 %	un numéro : 2 %
		aucun numéro : 6 %	un numéro : 6 %
Autres Etablissements	11 %	aucun numéro : 6 %	un numéro : 6 %
		aucun numéro : 73 %	un numéro : 27 %
Ensemble	100%	aucun numéro : 73 %	un numéro : 27 %

Source : Base de sondage de Génération 2004

Ce mauvais taux de remplissage s'explique par deux phénomènes très distincts. Tout d'abord sur le champ rectoral, on observe une absence totale de coordonnées téléphoniques. Ces dernières n'ont en effet pas pu être collectées du tout car absentes de la base centralisée. Or, le rectorat représente à lui seul plus de la moitié des sortants de la base de sondage.

Sur les autres champs, la variable est partiellement remplie : fichiers de mauvaise qualité ou refus de transférer des coordonnées téléphoniques expliquent ces valeurs manquantes.

Ainsi, on constate que notre enquête téléphonique va devoir mettre en œuvre des recherches de coordonnées téléphoniques massives. Pour cela, on a besoin de coordonnées postales bien remplies et fiables. Heureusement, 98,8 % des individus de la base de sondage possèdent des éléments d'adresse postale, ce qui constitue la base des recherches téléphoniques.

### 2.2.2. La mobilité géographique des individus

Pour autant, on peut posséder des coordonnées téléphoniques et/ou postale dans la base de sondage mais rester confronté au problème de la mobilité géographique des individus qui entame la fraîcheur des informations collectées. La question de la mobilité géographique est d'autant plus pertinente et

problématique pour les enquêtes Générations que les individus enquêtés sortent juste du système éducatif pour s'insérer sur le marché du travail. Le début de la vie active est propice à deux types de mobilité géographiques :

- Les mobilités géographiques pour motif professionnel
- Les mobilités géographiques pour cause de décohabitation

On constate que 20 % des jeunes de la Génération 2004 ont quitté leur région durant leurs études mais surtout 29 % d'entre eux ont travaillé hors de leur région de formation au cours des trois premières années de vie active. Ce taux dissimule pourtant de fortes disparités : plus les individus sont diplômés plus cette mobilité professionnelle est marquée.

**Tableau 3 : Les mobilités professionnelles de la Génération 2004**

	<b>Durant leurs études, ils ont quitté leur région (%)</b>	<b>Durant leurs trois premières années de vie active, ils ont travaillé hors de leur région de formation (%)</b>
Non diplômés	7	19
CAP ou BEP	9	20
Baccalauréat	16	25
Bac + 2	23	33
Licence	30	35
Bac + 4	36	39
Master	51	57
Doctorat	46	47
<b>Ensemble</b>	<b>20</b>	<b>29</b>

Source : Génération 2004

Champ : Jeunes ayant occupé au moins 1 emploi sur les 3 premières années de vie active pour la colonne 2.

Enfin, on sait qu'au moment où ils terminent leurs études, les jeunes résident majoritairement chez leurs parents : c'est le cas pour 74 % des hommes et 62 % des femmes de la Génération 2004. Trois ans après la fin de leurs études, ils sont 52 % chez les hommes et 31 % chez les femmes à habiter chez leurs parents. On observe encore une fois de fortes différences selon le niveau de diplôme de sortie.

**Tableau 4 : La décohabitation chez les jeunes de la Génération 2004**

	Ils habitaient chez leurs parents en 2004		Ils habitent chez leurs parents en 2007	
	Hommes	Femmes	Hommes	Femmes
	en %		en %	
Non diplômés	91	84	76	56
CAP-BEP-MC	86	78	67	42
Bac	79	72	56	39
Bac + 2	71	57	37	23
L3	52	39	28	17
M1	42	41	23	17
M2	36	33	11	9
Doctorat	9	10	5	4
<b>Ensemble</b>	<b>74</b>	<b>62</b>	<b>52</b>	<b>31</b>

Source : Génération 2004

Finalement, notre enquête téléphonique repose sur des coordonnées téléphoniques incomplètes et dont la fraîcheur pose problème. On comprend dès lors l'extrême importance des procédures d'enrichissement des coordonnées téléphoniques.

### 3. Le protocole d'enrichissement des coordonnées téléphoniques

#### 3.1. L'enrichissement des coordonnées téléphoniques dans l'enquête « Génération 98 »

La démarche de recherche des coordonnées téléphoniques dans les fichiers de France Télécom se décomposait en deux étapes successives :

- Des premières recherches sur l'adresse. Il s'agit bien de rechercher les individus grâce à leur nom, prénom et adresse précise. Le sous-traitant qui réalise l'opération nous retourne alors un code de fiabilité permettant de savoir si le numéro de téléphone a de fortes chances d'appartenir à l'individu (ou à ses parents). Pour les individus non retrouvés ou possédant un code de fiabilité jugé insuffisant des recherches élargies ont été entreprises.
- Puis des recherches élargies. Pour les individus ne possédant pas un code de fiabilité suffisant, une recherche élargie sur la France entière a été effectuée ; donc uniquement sur le nom et le prénom de l'individu. Cette recherche a souvent abouti à obtenir plusieurs numéros de téléphone d'homonymes parmi lesquels le numéro de téléphone de l'individu était susceptible de se trouver.

**Tableau 5 : Les résultats des recherches de coordonnées sur « Génération 98 »**

Individus retrouvés sur l'adresse	Nom+Prénom+ Adresse précise	1,5%
	Nom+Adresse précise	32,3%
	Nom+Tel	4,0%
	Prénom+Adresse précise	0,0%
	Adresse précise	0,6%
	Nom+Prénom+Adresse approchée	7,1%
Individus ayant des échos limités en recherche élargie	Nom+Prénom+Dép/Région/IDF/France	9,3%
Individus non retrouvés mais avec téléphone base de sondage		6,4%
Individus sans aucune coordonnées		38,8%
<b>Ensemble</b>		<b>100,0 %</b>

Source : Génération 1998

Au maximum 30 numéros de téléphone ont été conservés par le sous-traitant qui les a classés par ordre de priorité selon les niveaux d'élargissements suivants :

- le département de l'adresse initiale du fichier,
- la région de l'adresse,
- l'Ile de France,
- la France.
- 

Quelques remarques sur ces recherches :

- recherches automatiques sur nom et adresse : ces recherches représentent plus de la moitié des téléphones retrouvés. C'est la source principale des coordonnées téléphoniques disponibles. Et plus précisément c'est le téléphone des parents qui est retrouvé.
- recherches élargies : les 42 014 individus « retrouvés » en recherches élargies représentent 207 034 numéros de téléphone, soit une moyenne de 4,9 numéros de téléphone par individu (sachant qu'un maximum a été fixé à 30 numéros de téléphone).

### 3.2. Une expérience niçoise

Le Céreq a réalisé une extension d'échantillon sur Génération 98 pour le compte de l'université de Nice. Cependant, l'attrition de l'échantillon liée à la non-réponse ne satisfaisait pas les interlocuteurs niçois qui espéraient plus de questionnaires réalisés au vu du nombre d'individus présents dans la base de sondage. Il a donc été décidé d'un commun accord de tenter une expérimentation de recherches « artisanales » de coordonnées téléphoniques sur les individus non joints à l'issue du plateau d'enquête.

Cette opération a été réalisée par l'université en aval du plateau de l'enquête de Génération 98. Les enquêteurs ont effectué une réelle enquête de voisinage entièrement manuelle et les résultats sont intéressants. Sur 1 027 individus classés en faux numéros, en non joint ou en hors cible (mauvais individu au bout du fil) par le Céreq à l'issue du plateau, l'université de Nice a pu remonter les informations suivantes :

**Tableau 6 : Les résultats des recherches de Nice**

Dans le champ	404
Hors du champ	167
Injection d'un numéro à tester	34
Refus	18
Décédé	3
Non joint	401
<b>Ensemble</b>	<b>1 027</b>

Source : Génération 1998

Un des enseignements de ces recherches se trouve dans la répartition des types de recherches ayant abouti à retrouver ces 404 individus dans le champ. 30 % des individus retrouvés dans le champ l'ont été suite à une recherche nom+commune et 12 % sur une recherche nom+département. Au final, ces recherches ont permis de réaliser 236 enquêtes supplémentaires.

On peut doré et déjà constater sur ces effectifs niçois que les individus retrouvés suite à des recherches manuelles poussées et ceux contactés facilement sur le premier plateau d'enquête présentent des profils socioprofessionnels différents.

**Tableau 7 : Situation vis-à-vis de l'emploi des jeunes de la Génération 98 selon l'origine des données individuelles**

Situation à la date de l'enquête	Echantillon retrouvé par Nice	Echantillon retrouvé par le Céreq
Emploi	86 %	84 %
Chômage	11 %	8 %
Inactivité	1 %	3 %
Formation	1 %	2 %
Etudes	1 %	3 %

Source : Génération 1998

**Tableau 8 : Situation d'habitat des jeunes de la Génération 98 au 1<sup>er</sup> mars 2001 selon l'origine des données individuelles**

Situation à la date de l'enquête	Echantillon retrouvé par Nice	Echantillon retrouvé par le Céreq
En couple	54 %	38 %
Chez les parents	12 %	30 %
Seul	34 %	32 %

Source : Génération 1998

Ces différences portent sur de petits échantillons et sont donc à interpréter avec prudence.

On note de l'expérience niçoise que pour ces sortants de l'université, les numéros remontés, qui sont ceux de membres de la famille, sont bien des relais permettant de joindre l'individu majoritairement en couple ou seul sur un autre numéro.

### **3.3. Le protocole de Génération 2004**

Afin de profiter de l'expérience de Nice et d'améliorer le taux de réponse à l'enquête, le protocole de recherches de coordonnées téléphoniques a été rénové pour l'enquête Génération 2004. Les recherches téléphoniques dans le fichier France Télécom se sont ainsi déroulées en 2 phases successives :

- Recherches dites « A » : on effectue des recherches sur le nom, le prénom et l'adresse avec la possibilité d'avoir un libellé d'adresse approché ou un écho nom et adresse exacts mais un autre prénom (il s'agit alors des parents).
- Recherches dites « B et C » : ces recherches sont réalisées sur les rejets de la phase « A ».

- La phase B consiste à rechercher les individus en utilisant le nom/commune et le nom/département. Cette phase est nouvelle par rapport à Génération 98 et doit permettre de retrouver des membres de l'entourage familial du jeune. Grâce à l'intermédiaire de ce relais familial, on souhaite récupérer soit directement les coordonnées du jeune, soit un autre relais plus proche du jeune.

- La phase C est une recherche sur le Nom/Prénom/Commune et/ou sur le Nom/Prénom/département. On espère par cette recherche homonymique retrouver des jeunes qui auraient connu une mobilité géographique dans un périmètre proche de leur commune de résidence en 2004. On s'en tient à ce périmètre géographique car les recherches de Génération 98 ont fait apparaître une rentabilité faible des recherches plus étendues.

A l'issue de ces recherches, voilà comment se répartissaient nos 371 600 individus de l'échantillon.

**Tableau 9 : Les résultats des recherches de coordonnées sur « Génération 2004 »**

Résultat des recherches	Effectif	Pourcentage
Retrouvé en phase A	134 530	36 %
Retrouvé en phase B uniquement	99 350	27 %
Retrouvé en phase C uniquement	9 910	3 %
Retrouvé en phase B et C	6 070	2 %
Numéro Base de sondage uniquement	37 910	10 %
Aucun numéro à l'issue des recherches	83 830	22 %
Ensemble des individus	371 600	100 %

Source : Génération 2004

Au final, deux modifications substantielles sont à retenir dans le protocole de recherche des coordonnées téléphoniques.

- suppression de recherches Nom+Prénom+Département/Région/IDF/France qui représentaient plus de 9 % des échos. Ces recherches ont été jugées peu rentables.
- Ajout de recherches Nom+Commune/Département qui représentent sur Génération 2004 un gain net de 27 % d'échos en recherches.

Ces modifications expliquent en quasi totalité la baisse du nombre d'individus non enrichis par les recherches et n'ayant aucun numéro dans la base de sondage. Ces individus qui sont inutilisables pour notre enquête passent ainsi de 39 % à 22 % de l'échantillon entre Génération 98 et Génération 2004. Ces 22 % d'individus représentent toujours une perte nette de taux de réponse à notre enquête.

Toutefois, ce gain de 17 points sur les individus non utilisables n'implique pas directement, et c'est bien dommage, un gain de taux de réponse proportionnel. Il convient, en effet, de mesurer la « rentabilité » des coordonnées ainsi obtenues.

Ces numéros injectés sur notre plateau d'enquête correspondent théoriquement à des coordonnées de relais familiaux. Il nous a donc fallu préparer un questionnaire sous CATI propre à ce contact particulier. Le dialogue, lors de ce contact avec un parent parfois éloigné, devait être totalement revu pour permettre la récupération efficace des coordonnées du jeune.

## 4. Quels résultats pour quel taux de réponse ?

### 4.1. Un gain net de répondants

Les résultats d'appels issus du plateau de Génération 2004 nous permettent de calculer le taux de réponse à l'enquête suivant :

$$\frac{\text{Nb individus répondants} + \text{Nb individus répondants Hors Champ}}{\text{Nb individus échantillonnés}} = \frac{65\,361 + 50\,355}{371\,497} = 31,2 \%$$

Sur les 99 350 individus qui ont bénéficié, comme type unique d'écho, d'un et de plusieurs retours sur le nom+commune/département à l'issue des recherches téléphoniques, 68 023 ne possédaient pas de téléphone base de sondage. Ces individus représentent donc un gain net puisque sans ces échos issus des nouvelles recherches, ils n'auraient pas pu être mis en production sur le plateau d'enquête.

Examinons quels ont été les résultats d'appel de ces individus.

**Tableau 10 : Résultat d'appel des individus selon l'enrichissement**

	Individus enrichis sur nom+commune/département		Ensemble des individus envoyés sur les plateaux d'enquêtes	
Hors Champs	5 187	7,6 %	50 355	17,5 %
Hors Cible	27 561	40,5 %	50 207	17,5 %
Enquête réalisée	5 911	8,7 %	65 361	22,7 %
Refus répondre de l'individu + tiers	12 637	18,6 %	40 253	14,0 %
Autre non réponse*	16 727	24,6 %	81 495	28,3 %
Ensemble	68 023	100,0 %	287 671	100,0 %

\* Autre non réponse : réside à l'étranger, injoignable permanent, a déjà répondu, faux numéro, ne répond pas, à rappeler, occupé, répondeur

Source : Génération 2004  
Fichier historique des Appels

Le premier constat qui apparaît à la lecture de ce tableau est la proportion élevée d'individus issus des nouvelles recherches qui finissent classés en « Hors Cible ». Ce constat est finalement assez logique si l'on considère le fait que les individus enrichis sur le nom et la commune/département renvoient fréquemment sur des homonymes noms. De même, les refus de répondre s'expliquent par des refus de tiers élevés liés au fait que les numéros enrichis permettent de joindre principalement des relais familiaux et non pas l'individu lui-même, auquel cas le prénom serait validé aussi. D'ailleurs, un rapide test du chi-deux entre la réponse à l'enquête et le fait de provenir de ces nouveaux enrichissements conduit à rejeter l'hypothèse d'indépendance des deux phénomènes.

Si les résultats soulignent un plus faible rendement des individus enrichis sur le nom et la commune/département, il convient toutefois d'insister que ces enrichissements ont permis d'interroger plus de 11 000 individus dont 5 187 se sont avérés hors champs et 5 911 ont pu répondre à la totalité du questionnaire. Ces 11 000 répondants représentent un gain net lié à ces nouvelles recherches. On peut ainsi simuler, qu'en l'absence de ces enrichissements, nous n'aurions obtenu que 104 618 répondants soit un taux de réponse de 28,2 %.

#### **4.2. Des profils d'individus qui diffèrent**

On constate que les répondants au questionnaire issus des nouveaux enrichissements présentent des caractéristiques qui se détachent de celles de notre population. Si l'on s'intéresse au mode d'habitat de ces individus, on note qu'ils résident plus souvent chez leurs parents que les autres répondants à l'enquête. Cela peut s'expliquer par le fait que ces enrichissements permettent souvent de retrouver la famille du jeune et donc de joindre des individus qui résident encore chez leurs parents (au sens large). Ce constat résiste à l'examen des intervalles de confiance puisque les valeurs des individus issus des enrichissements nom+commune/département se détachent nettement de celles des autres répondants.

On peut souligner les différences de résultats observées entre les recherches niçoises sur Génération 98 qui se limitaient aux sortants de l'université et retrouvaient des individus vivant seuls ou en couple et

les recherches similaires menées sur Génération 2004 qui retrouvent des jeunes résident plus souvent chez leurs parents. En effet, ces jeunes sont plus fréquemment des sortants du secondaire.

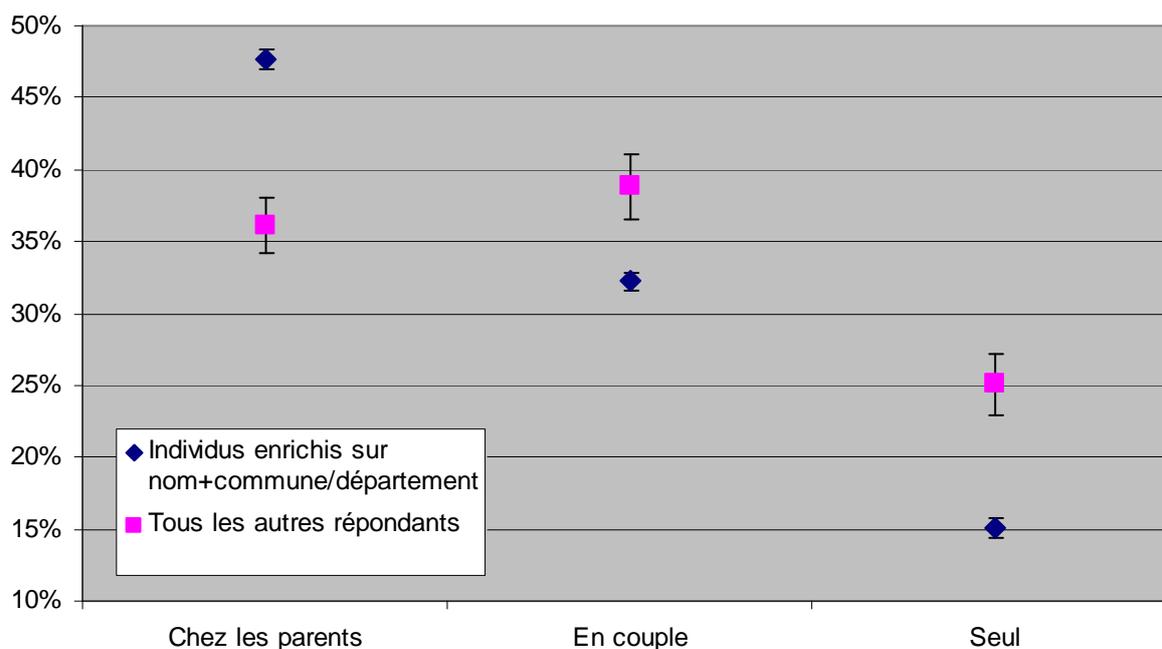
**Tableau 11 : Répartition des individus selon le type d'enrichissement et le mode d'habitat**

	Situation d'habitat à la date d'enquête			Ensemble
	En couple	Chez les parents	Seul	
Individus enrichis sur nom+commune/département	1 705 32,2 %	2 525 47,7 %	1 066 15,1 %	5 296
Tous les autres répondants	20 990 38,8 %	19 528 36,1 %	13 606 25,1 %	54 124
Ensemble	22 695	22 053	14 672	59 420*

Source : Génération 2004

\* Le total de 59 420 diffère du nombre d'enquêtes réalisées du tableau 10 car il s'agit de questionnaires apurés et exploitables.

**Graphique 2 : Répartition des individus selon le type d'enrichissement et le mode d'habitat avec Intervalle de confiance**



Source : Génération 2004

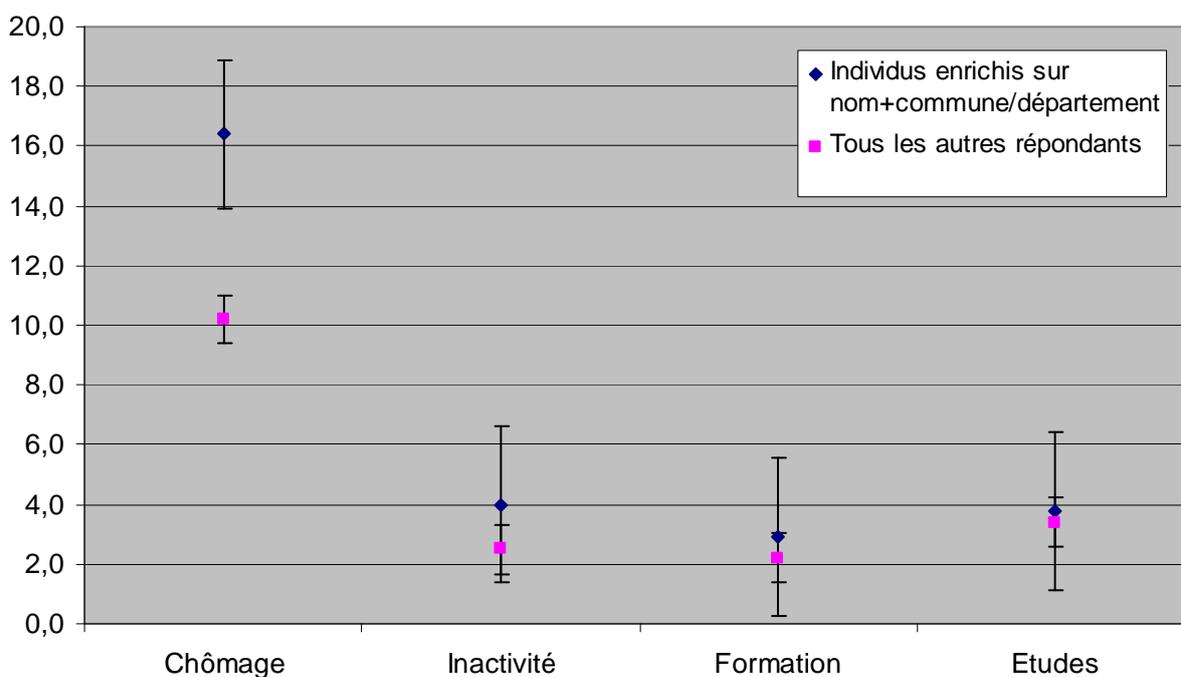
Si le fait de retrouver des individus habitant plus souvent chez leurs parents n'est pas du tout étonnant vue la méthode de recherches des coordonnées, il est plus intéressant de constater que la différenciation des individus se poursuit si l'on examine la situation vis-à-vis de l'emploi.

**Tableau 12 : Répartition des individus selon le type d'enrichissement et la situation vis-à-vis de l'emploi**

	Situation d'emploi à la date d'enquête					Ensemble
	Emploi	Chômage	Inactivité	Formation	Etudes	
Individus enrichis sur nom+commune/département	3 861 72,9 %	869 16,4 %	215 4,0 %	152 2,9 %	199 3,8 %	5 296
Tous les autres répondants	44 230 81,7 %	5 517 10,2 %	1 322 2,5 %	1 230 2,2 %	1 825 3,4 %	54 124
Ensemble	48 091	6 386	1 537	1 382	2 024	59 420

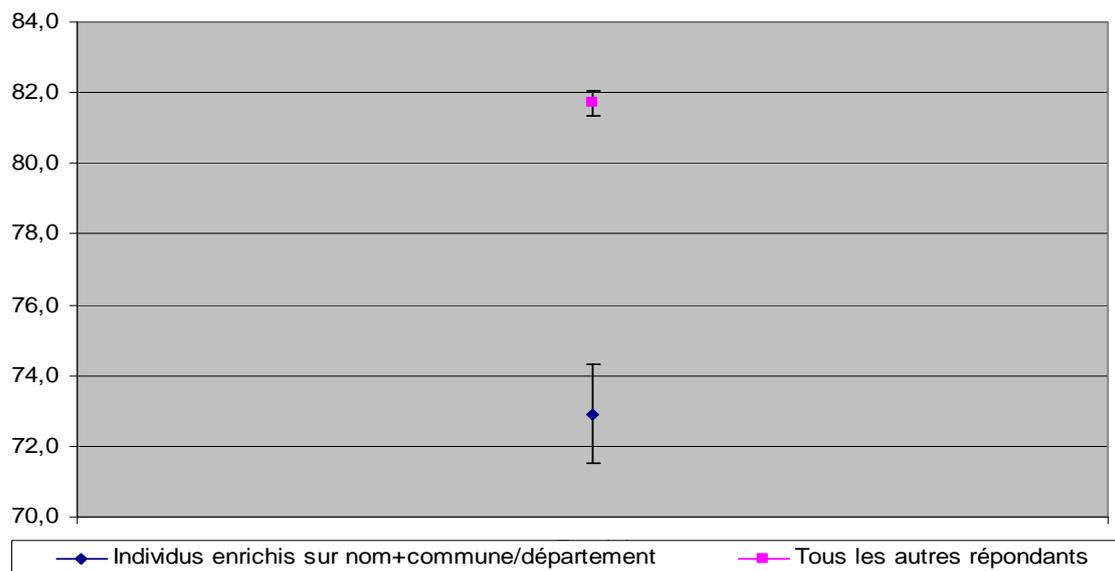
Source : Génération 2004

**Graphique 3 : Répartition des individus selon le type d'enrichissement et les situations d'inactivité et de chômage avec Intervalle de confiance**



Source : Génération 2004

**Graphique 4 : Répartition des individus en emploi selon le type d'enrichissement  
avec Intervalle de confiance**



Source : Génération 2004

Concernant la répartition des individus en emploi ou non, l'impact des recherches de coordonnées téléphoniques est assez net : les individus retrouvés à l'issue de ces nouvelles recherches occupent moins souvent des emplois que les individus issus d'autres enrichissements. On peut compléter ce constat en précisant que les individus issus des nouvelles recherches sont plus souvent au chômage que les autres jeunes. Pour autant, et à cause du chevauchement des intervalles de confiance, il est impossible de préciser des différences sur les situations d'inactivité.

Un dernier tableau doit nous permettre d'achever ce constat.

**Tableau 13 : Répartition des individus selon le type d'enrichissement et le niveau de sortie des études**

	Niveau de sortie des études*								Ensemble
	I	II	III	IV	IV+	V	Vb	VI	
Individus enrichis sur nom+commune/département	279 5,3 %	269 5,1 %	747 14,1 %	1 508 28,5 %	382 7,2 %	1 667 31,4 %	279 5,3 %	165 3,1 %	5 296
Tous les autres répondants	10 457 19,3 %	10 452 19,3 %	8 705 16,1 %	8 561 15,8 %	4 982 9,2 %	9 440 17,4 %	1 022 1,9 %	505 1,0 %	54 124
Ensemble	10 736	10 721	9 452	10 069	5 364	11 107	1 301	670	59 420

Source : Génération 2004

\* Niveaux de formation :

I : Troisième cycle universitaire ou grande école diplômé

II : Second Cycle diplômé

III : Bac+2 diplômé

IV : Classe de terminale diplômé ou non du BAC

IV+ : Diplômé du Bac ayant poursuivi des études dans le supérieur mais pas de diplôme bac+2

V : année terminale des cycles courts professionnels et abandons de la scolarité du second cycle long avant la classe terminale

Vb : 3ème générale, de 4ème et 3ème technologiques et classes du second cycle court avant l'année terminale

VI : 1er cycle du second degré (6ème, 5ème, 4ème) et formations préprofessionnelles en un an

La comparaison des deux populations par niveau de sortie laisse clairement apparaître que les nouvelles recherches permettent de joindre plus de sortants des niveaux IV et peinent à fournir un contact auprès des jeunes les plus diplômés. Quoique les effectifs ne permettent pas de trancher de façon absolue, il semble enfin que les enrichissements de type nom+commune/département fassent remonter plus de questionnaires auprès des jeunes sans aucun diplôme.

### 4.3. Quel redressement ?

Ces individus récupérés grâce aux nouvelles techniques d'enrichissement présentent certes des profils différents ; mais ils viennent surtout nourrir des catégories de jeunes que l'enquête peine à collecter. Ainsi, les répondants ayant des faibles niveaux de diplômes sont sous-représentés dans les enquêtes Générations. C'est ce qu'illustre le tableau ci-dessous. Après redressement de la non-réponse et avant calage sur l'Enquête Emploi, les enquêtes Générations présentent des déficits marqués sur les plus bas niveaux de qualification. Inversement, les sortants des plus hauts niveaux de diplômes sont souvent surreprésentés.

**Tableau 14 : Répartition des répondants aux enquêtes Générations par niveau de sortie avant et après calage de la pondération**

Niveau de sortie	G98		G01		G04	
	Non calée	Calée	Non calée	Calée	Non calée	Calée
1 et 2	22,1	18,6	26,7	19,3	26,5	23,4
3	18,4	18,6	17,7	17,9	16	17,3
4	19,9	17,4	17,9	16	19,8	19,3
4+	11,2	12,3	13,1	11,1	12,3	11,1
5	22,2	25,2	20,3	27,9	21,5	22
5b et 6	6,2	7,9	4,3	7,8	4	6,9

Source : Génération 2004

Il est ainsi intéressant de constater que, si l'on retire ces individus issus des recherches nom+commune/département, ce décalage entre les données non calées et les données calées s'accroît encore, notamment sur les niveaux extrêmes.

**Tableau 15 : Répartition des répondants à Génération 2004 selon le type d'enrichissement et le niveau de sortie avant et après calage de la pondération**

Niveau de sortie	G04		
	Non calée sans nouvelles recherches	Non calée tous individus	Calée tous individus
1 et 2	31,83	26,5	23,4
3	16,8	16	17,3
4	17,4	19,8	19,3
4+	13,5	12,3	11,1
5	17,6	21,5	22
5b et 6	2,8	4	6,9

Source : Génération 2004

Ainsi, outre le gain de taux de réponse, ces individus récupérés grâce aux enrichissements nom+commune/département permettent de réduire la variance des poids après calage.

**Tableau 16: Taux de chômage des individus dans Génération 2004 et dans l'Enquête Emploi**

Taux chômage des Individus issus des enrichissements nom+commune/département		Taux de chômage des autres Individus		Taux de chômage EE 2007 – trimestre 2	
Non pondéré	Pondéré	Non pondéré	Pondéré	Ancienne interprétation	Nouvelle interprétation
18,4 %	22,1 %	11,1 %	12,4 %	12,8%	12,2%

Sources : Enquête Génération 2004  
Enquête Emploi 2007 – trimestre 2 – sur les sortants 2004

#### **4.4. Un coût important**

Le calcul du prix de ces questionnaires gagnés à l'issue des recherches sur le nom + commune/département constitue une étape importante du bilan de l'opération. Une analyse coût/avantage de la méthode passe ainsi par cette estimation du surcoût financier lié à la nouvelle méthodologie.

Un premier exercice permet rapidement de faire apparaître 2 types de coûts qui se succèdent pour la réalisation des questionnaires :

- a. le prix des contacts téléphoniques dits « qualifiés ». Les contacts qualifiés regroupent tous les contacts téléphoniques qui ont abouti à retrouver le bon individu ainsi que la non-réponse autre que : répondeur, ne répond pas, à rappeler, occupé, faux numéro disque France Télécom. Il s'agit uniquement des contacts dits « qualifiés » du plateau de pré qualification (voire explication ci-dessous)
- b. le prix de réalisation de l'enquête téléphonique elle-même.

Par exemple, un individu peut disposer de 3 numéros de téléphones dans notre base. On peut ainsi avoir l'enchaînement d'appels suivants :

- Ne répond pas (numéro1)
- Occupé (numéro1)
- Hors Cible (numéro2) -> contact qualifié
- Ne répond pas (numéro3)
- Enquête réalisée (numéro3)

La société d'appel nous facturera alors 1 contact qualifié et 1 enquête réalisée pour le même individu. Précisons que pour les contacts qualifiés comme pour les enquêtes réalisées, la facturation diffère légèrement selon si l'appel a été réalisé sur un numéro de fixe ou sur un numéro de portable ou numéro spécial.

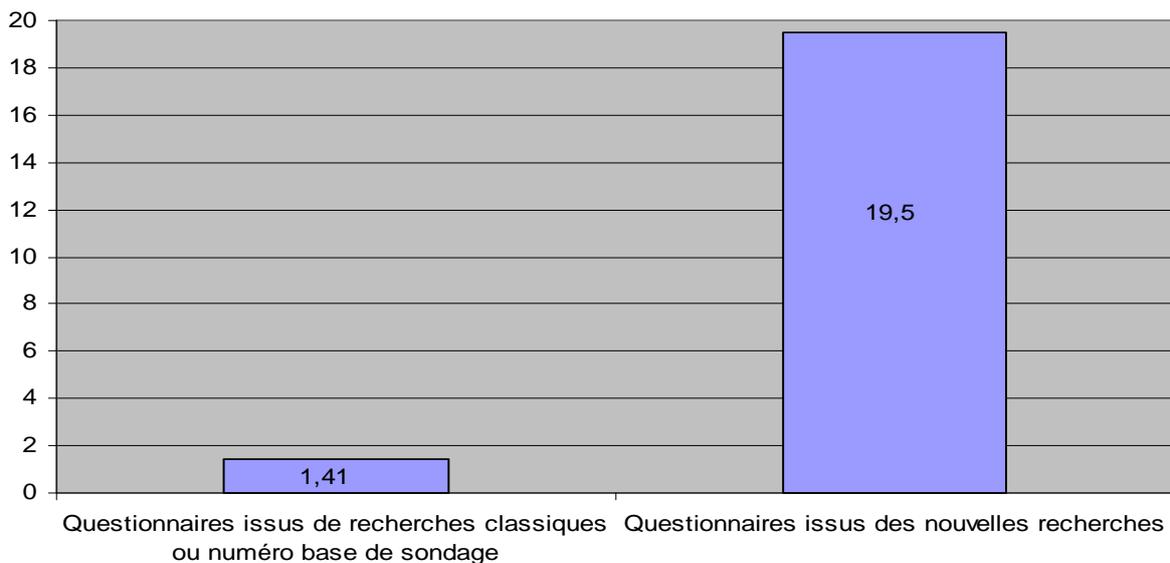
### a. Les contacts téléphoniques qualifiés

Le plateau de pré qualification est un plateau qui permet de s'assurer que certains numéros de téléphones issus des recherches B et C appartiennent bien au bon individu avant de passer au plateau d'enquête lui-même.

Au total, la société d'appel a qualifié 132 530 contacts en plateau de pré qualification. Sur ces 132 530 contacts qualifiés, 57,5% concernent des individus ayant été enrichis uniquement suite à une recherche nom+commune/département. Si l'on tient compte de la proportion de téléphones fixes et portable, la facture de la qualification pour les individus issus des nouvelles recherches s'élève donc à 103 400 euros. Rapporté aux 5 300 questionnaires exploitables (après apurement), cela entraîne un coût unitaire de la qualification pour ces individus de 19€50 par questionnaire.

En comparaison, le même coût par questionnaire sur les autres individus représente 1€41. Le surcoût lié à la qualification de ces questionnaires issus des nouvelles recherches est impressionnant.

**Graphique 5 : Différence de coût liée à la qualification selon le type de recherches**



Source : Génération 2004

La différence de coût entre les questionnaires issus des nouvelles recherches et les autres questionnaires se décompose en 2 volets

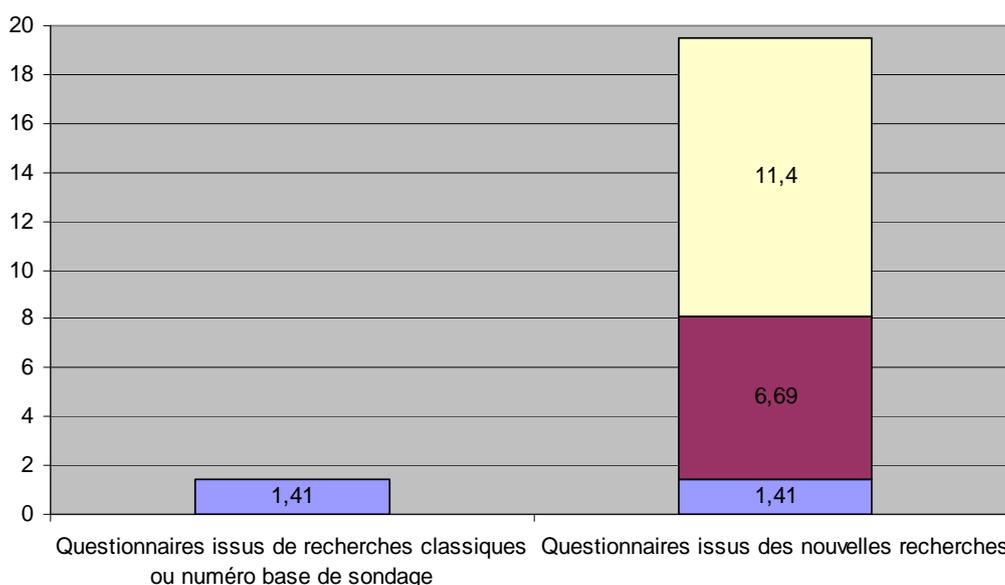
- les individus issus de ces nouvelles recherches ont un taux de réponse beaucoup plus faible que les autres. On a notamment un nombre important de contacts qui aboutissent en Hors Cible (mauvais individus sur le numéro) et donc par conséquent une proportion faible d'enquêtes réalisées. Ainsi, si l'on fait supporter le coût de la qualification non pas sur les 5 300 questionnaires effectivement obtenus mais sur le nombre de questionnaires qu'aurait générés ces recherches si le taux de questionnaires réalisés avait été le même que sur l'ensemble des individus, on explique un écart de 11€40 par questionnaire.

- le nombre moyen de contacts qualifiés est plus élevé chez les individus issus des nouvelles recherches que sur les autres individus. Un calcul du rapport de vraisemblance nous permet d'affirmer que les individus non issus des nouvelles recherches ont deux fois plus de chance de n'avoir qu'un seul contact qualifié que les individus issus des nouvelles recherches. En moyenne, les individus issus des nouvelles recherches ont abouti en moyenne à 1,76 contacts qualifiés contre 1,41 pour les autres. Un test d'égalité des moyennes nous amène à rejeter l'égalité de ces deux valeurs : l'écart est donc significatif.

Toutefois, on est tenté de s'interroger : la mesure de cet écart résiste-t-elle à un raisonnement toutes choses égales par ailleurs, notamment si l'on considère la structure des données par source (rectorale, université, etc. ...). Une régression linéaire par les MCO nous amène à constater que la mesure de cet écart varie peu à structure des données par source égale. L'influence du type de recherches téléphoniques est confirmée.

Cet écart explique ainsi une différence de 6,69 par questionnaire.

**Graphique 6 : Explication de la différence de coût lié à la qualification**



Source : Génération 2004

#### b. Coût des enquêtes réalisées

Ce deuxième type de coût implique par définition peu de différence entre les questionnaires issus de nouvelles recherches et les autres. Toutefois, et si l'on se souvient que le coût d'une enquête réalisée est différent selon si l'enquête est réalisée sur un numéro de fixe ou sur d'autres types de numéros, il s'agit de vérifier cette donnée.

En effet, la part des enquêtes réalisées sur portables et numéros spéciaux est de 47,8 % pour les enquêtes issues des nouvelles recherches contre 39,2 % pour les autres enquêtes. Les recherches élargies, qui permettent de récupérer des numéros fournis par des tiers et non présents dans les annuaires, induisent là encore un surcoût au questionnaire. Ainsi, les questionnaires d'enquêtes issus de ces recherches reviennent en moyenne à 10€49 contre 10€25 pour les autres questionnaires.

Bilan : Si l'on considère les coûts de qualification des contacts plus les coûts de passation des questionnaires, les 5 300 questionnaires gagnés grâce aux nouvelles recherches affichent un coût unitaire de 29€99 contre 11€66 pour tous les autres questionnaires ; ils sont 2,6 fois plus coûteux.

## Conclusion

A l'heure où les enquêtes téléphoniques affichent des baisses historiques de taux de réponses, les sondeurs rivalisent d'ingéniosité pour contrecarrer cette tendance. Précurseurs, les Etats-Unis affichent dans un rapport fréquemment cité<sup>4</sup> des taux de refus de répondre multipliés par deux en 15. Pressentant cette dégradation dans les enquêtes Génération, le Céreq a anticipé cette baisse des taux de réponse par la mise en œuvre d'une nouvelle méthode d'enrichissement des coordonnées téléphoniques.

L'enquête Génération 2004 s'offre ainsi le luxe d'une augmentation d'un point et demi de son taux de réponse par rapport à Génération 1998. On estime que les questionnaires gagnés par l'intermédiaire de cette nouvelle méthode représentent un gain de taux de réponse d'environ trois points. Et pour souligner encore les bénéfices de cette méthode, il convient d'insister sur le fait que ces répondants ainsi joints viennent étoffer des effectifs que les enquêtes Génération peinent habituellement à contacter (jeunes chômeurs sans diplôme et/ou sans qualification). Ce gain de questionnaires permet donc de réduire la variance des poids et d'améliorer la précision des estimateurs.

Pourtant, l'examen des coûts – avec des questionnaires deux fois et demie plus chers que les autres – ne nous permet pas de valider totalement la méthode. Des pistes d'amélioration doivent être envisagées dans d'autres directions, notamment en amont de la phase d'enrichissement. La collecte de la base de sondage de Génération 2007 entamée courant janvier 2009 offre ainsi de nouvelles perspectives. En effet, on a pu, pour la première fois, collecter des coordonnées téléphoniques pour près de deux tiers des jeunes des fichiers rectoraux. La collecte de l'information a également été enrichie sur le champ de l'apprentissage et des universités.

Enfin, la rentabilité en termes de nombres de questionnaires de cette méthode d'enrichissement souligne les limites d'une enquête de voisinage – telle que l'Université de Nice a pu la mettre en œuvre – à grande échelle. La nécessité de recourir à des systèmes CATI pour le plan de dialogue ainsi qu'à des prédictifs d'appels détourne l'intérêt initial de la méthode. Des pistes alternatives et sans doute plus réalistes sont à explorer dans le sens d'une meilleure collecte des données et dans la mise en place d'un annuaire téléphonique universel.

---

<sup>4</sup> R. Curtin, S. Presser et E. Singer, « *Changes in Telephone Survey nonresponse over the Past Quarter Century* », 2005.

## Bibliographie

- [1] P. Ambroise et P. Mauris, « L'usage du téléphone dans les sondages », 1999
- [2] F. Beck, S. Legleye et P. Peretti-Watel, « Aux abonnés absents : liste rouge et téléphone portable dans les enquêtes en population générale sur les drogues », 2005
- [3] JM. Brignier et F. Dupont, « Taux de réponse et qualité dans les enquêtes téléphoniques : les enseignements des études d'audience de la Presse Quotidienne d'IPSOS et de l'étude d'audience radio et cinéma de Médiamétrie », 2005
- [4] JM. Firdion, « L'effet du rang d'appel et de la présence du conjoint dans une enquête par téléphone », 1993
- [5] B. Riandey et JM. Firdion, « Vie personnelle et enquête téléphonique, l'exemple de l'enquête ACSF », 1993
- [6] G. Roy et A. Vanheuverzwyn, « Le téléphone mobile dans les enquêtes par sondage », 2002
- [7] P. Tassi, « La qualité des enquêtes téléphoniques, l'échantillon des répondants », 1992



## INDEX DES SIGLES

---

Agefiph	Association pour la gestion des fonds pour l'insertion professionnelle des personnes handicapées
Anfa	Association nationale pour la formation automobile
BCE	Banque centrale des établissements (répertoire d'établissements de formation géré par l'Education nationale)
BEP	Brevet d'études professionnelles
BTS	Brevet de technicien supérieur
CAP	Certificat d'aptitude professionnelle
CDES	Commissions départementales de l'éducation spéciale
Céreq	Centre d'études et de recherches sur les qualifications
CFA	Centre de formation d'apprentis
Cnil	Commission nationale Informatique et Liberté
Cnis	Conseil national de l'information statistique
CNU	Conseil national des universités
Cotorep	Commission technique d'orientation et de reclassement professionnel
Dares	Direction de l'animation de la recherche, des études et des statistiques (service statistique ministériel du Travail)
Depp	Direction de l'évaluation, de la prospective et de la performance (service statistique ministériel de l'Education nationale)
Drees	Direction de la recherche, des études, de l'évaluation et des statistiques (service statistique ministériel de la Santé et des Affaires sociales)
DRJS	Direction régionale de la jeunesse et des sports
Finess	Fichier national des établissements sanitaires et sociaux (géré par la DREES)
Forco	Observatoire prospectif du commerce
GSE	Groupe statistique de l'évaluation (au sein du Conseil national de la formation tout au long de la vie)
Ifen	Institut français de l'environnement
INE	Identifiant national d'étudiant
Insee	Institut national de la statistique et des études économiques
Iufm	Institut universitaire de formation des maîtres
IUT	Institut universitaire de technologie
MEN	Ministère de l'éducation nationale
Nes	Nomenclature économique de synthèse (gérée par l'INSEE)
NSF	Nomenclature des spécialités de formation (gérée par l'Education nationale)
PCS	Profession et catégories socioprofessionnelle (nomenclature gérée par l'INSEE)

REFLET	Regards sur les flux de l'enseignement technique et professionnel (base de données statistiques gérée par le CEREQ)
Sicore	Système Informatique de codage des réponses aux enquêtes (INSEE)
Sise	Système d'information et de suivi des étudiants



ISSN 1776-3177  
Marseille, 2010.