

Article

Symposium 2008 :
Collecte des données : défis, réalisations et nouvelles orientations

Évaluation et traitement de la non-réponse dans la cohorte Elfe : résultats des études pilotes

par Hélène Sarter, Nicolas Razafindratsima et l'équipe Elfe
2009



Évaluation et traitement de la non-réponse dans la cohorte Elfe : résultats des études pilotes

Hélène Sarter, Nicolas Razafindratsima et l'équipe Elfe ¹

Résumé

La non-réponse est inévitable dans toute enquête, malgré tous les efforts mis en œuvre pour la réduire aux différentes étapes de l'enquête. Cette non-réponse peut notamment être responsable de biais dans l'estimation. La non-réponse est par ailleurs une problématique d'autant plus importante dans une étude longitudinale que l'échantillon se réduit au fil du temps. L'Étude Longitudinale Française depuis l'Enfance (Elfe) est un projet de suivi de 20 000 enfants de la naissance à l'âge adulte dans une approche pluridisciplinaire. Cet article est basé sur les résultats des premières études pilotes menées en 2007 qui ont permis de tester la faisabilité et l'acceptation de l'étude. Les résultats de participation sont présentés (taux de réponse, facteurs liés à la non-réponse) ainsi qu'une première ébauche des méthodes de traitement de la non-réponse envisagées.

Mots clés : Non-réponse, taux de réponse, cohorte, enquête longitudinale, pondération longitudinale.

1. Introduction

Toute enquête est confrontée au problème de la non-réponse, dans une plus ou moins grande mesure, qu'il s'agisse de la non-réponse totale (absence totale de réponse) ou partielle (non-réponse à certains items seulement). Par ailleurs, au cours des dernières décennies, un déclin dans les taux de réponse à été observé dans de nombreuses études (De Leeuw et De Heer, 2002) : il est de plus en plus difficile de contacter les individus mais aussi de les persuader de participer. La non-réponse est responsable de plusieurs types de difficultés dans l'analyse des données.

Tout d'abord, la réduction de la taille de l'échantillon entraîne une perte de puissance statistique et donc une augmentation de la variance des estimations. Ensuite, la non-réponse peut introduire un biais dans les estimations portant sur des variables d'intérêt : les individus non-répondants sont en effet susceptibles de différer des individus répondants en termes de caractéristiques directement ou indirectement liées aux variables d'intérêt. Enfin, la non-réponse complique les méthodes d'estimation de la variance.

La première chose à faire est de limiter la non-réponse en mettant en place des mesures efficaces d'incitation à participer, des relances et des méthodes de suivi des adresses par exemple, en limitant le fardeau de non-réponse, etc. Une fois les données collectées, la non-réponse peut être traitée par des méthodes statistiques visant à réduire le biais dû à la non-réponse, notamment les méthodes de repondération (on augmente les poids de sondage des répondants pour compenser la non-réponse) et les méthodes d'imputation (on remplace la valeur manquante par une valeur choisie selon une méthode donnée et menant à une valeur la plus plausible possible). D'une manière générale, les méthodes de repondération sont préférentiellement utilisées pour le traitement de la non-réponse totale alors que les méthodes d'imputation sont plus adaptées pour le traitement de la non-réponse partielle.

Le problème de la non-réponse est par ailleurs d'autant plus important dans une étude longitudinale que l'échantillon se réduit au fil du temps (attrition ou érosion). Les études longitudinales sont également confrontées à

¹Hélène Sarter, Institut de Veille Sanitaire (InVS), Département Santé-Environnement, 12, rue du Val d'Osne, 94 415 Saint-Maurice Cedex, France (h.sarter@invs.sante.fr); Nicolas Razafindratsima, Institut National d'Études Démographiques (INED), 133 boulevard Davout, 75980 Paris Cedex 20, France (razafind@ined.fr); l'équipe Elfe : Henri Leridon (Directeur), Corinne Bois, Claudine Pirus et Stéphanie Vandentorren (coordinatrices), Pascal Arduin, Arnaud Dorléans, Amivi Oleko, Ruxandra Popa, Karen Touboulic.

la non-réponse de vague qui intervient lorsqu'un individu ne répond pas à une ou plusieurs vagues. Le présent article est basé sur les premiers résultats des études pilotes du projet d'étude longitudinale Elfe.

2. Contexte : le projet de cohorte Elfe

2.1 Description générale

L'Étude Longitudinale Française depuis l'Enfance (Elfe) est un projet de cohorte représentative au plan national de 20 000 enfants qui seront suivis de la naissance à l'âge adulte (www.elfe2009.fr). Il s'agira de la première cohorte de naissances d'une telle ampleur en France. L'étude permettra d'analyser le développement de l'enfant autour de trois grands axes : la santé de l'enfant, les relations entre environnement et santé et les aspects sociodémographiques. Il s'agit donc d'un projet très largement pluridisciplinaire auquel participent à l'heure actuelle environ 60 équipes de recherche pour environ 90 projets de recherche spécifiques. Les enfants seront inclus dès la naissance en maternité et seront ceux nés pendant 16 jours donnés d'une année. Pendant le suivi de la cohorte, il est prévu de rassembler des données recueillies par des collectes directes (questionnaires, prélèvements biologiques, carnets de suivi...) mais également par différentes sources externes qui seront couplées aux données recueillies (enquêtes médicales en milieu scolaire, suivi de scolarité, données de l'assurance maladie...) à des âges pertinents. Le démarrage est prévu en 2010-2011 en fonction de l'avancement du plan de financement du projet.

2.2 Unité enquêtée et champ de l'enquête

L'enquête vise à obtenir un échantillon représentatif des naissances vivantes d'une année donnée en France. Pour des raisons pratiques, les naissances multiples au-delà de deux ne seront pas incluses. L'unité statistique longitudinale est l'enfant. Dans un premier temps, la mère et/ou le père selon les enquêtes répondront aux questions jusqu'à ce que l'enfant soit en mesure de répondre également (âge à définir).

2.3 Description du sondage

Echantillon

Les enfants inclus seront ceux nés une année donnée dans toutes les maternités de France pendant les jours correspondants à l'Échantillon Démographique Permanent (EDP) rénové de l'Insee, soit 16 jours répartis en 4 périodes de 4 jours dans l'année (à chaque trimestre). La base est donc proche d'un échantillon au 1/25^{ème}. A l'heure actuelle, l'EDP rassemble tous les actes d'état civil et toutes les informations consignées dans les bulletins de recensement successifs relatives aux personnes nées l'un des jours de référence d'une année quelconque. L'échantillon ainsi constitué sera représentatif des naissances de l'année et permettra ensuite un repérage facile de ces enfants dans d'autres sources statistiques par leur date de naissance.

Nombre de sujets

Le nombre de naissances en France en 2006 était d'environ 830 000 (Richet-Mastain, 2007), ce qui nous assure d'obtenir un échantillon d'au moins 20 000 enfants inclus dans la cohorte avec un taux de réponse au moins égal à 55 %. Ce nombre de sujets permettra de répondre à l'ambition longitudinale et pluridisciplinaire du projet (compte tenu des taux d'attrition à prévoir), de réaliser des analyses par sous-groupes ou pour des caractéristiques rares. Par ailleurs, cet échantillon de 20 000 enfants permettra la mise en relation entre une exposition et un effet (exposition et effet au sens large, que ce soit dans le domaine de la santé ou en sciences sociales) pour des expositions et/ou effets rares.

2.3 Mode de collecte

Un certain nombre de rendez-vous et de modes de collecte sont déjà planifiés, notamment le recrutement en maternité et l'enquête réalisée lorsque l'enfant a 2 mois.

L'enquête en maternité :

L'inclusion sera assurée par des sages-femmes enquêtrices spécifiquement dédiées au projet dans les maternités. Ce mode d'inclusion en maternité implique une mobilisation des services de maternité pendant les périodes de recrutement dans l'ensemble des maternités de France, soit un peu moins de 600 maternités. Ceci nécessite le recrutement et la formation d'environ 1 200 personnes par collecte de 4 jours. Les solutions envisagées sont de recruter des sages-femmes des maternités elles-mêmes, des élèves sages-femmes et de jeunes retraitées. Une coordination régionale sera également mise en place afin de faciliter la coordination entre les différents acteurs et d'assurer une bonne formation des enquêtrices. Cette première étape en maternité comprend : un recueil d'informations du dossier médical et un questionnaire auprès de la mère sur le déroulement de la grossesse, la période périnatale et l'état de santé de femmes et des enfants à l'accouchement pour l'ensemble des naissances indépendamment de l'acceptation du suivi ; le consentement au suivi Elfe ; des prélèvements biologiques également soumis à consentement (sang de cordon, urines et cheveux de la mère, lait maternel) et un auto-questionnaire sur l'alimentation et le mode de vie pendant la grossesse pour les mères ayant accepté le suivi Elfe.

L'enquête à 2 mois

L'enquête sera réalisée par un enquêteur de l'Institut national de la statistique et des études économiques (Insee) sur la base d'un questionnaire assisté par ordinateur en face-à-face à domicile pour la mère (1 heure environ) et par téléphone pour le père (20-30 minutes). Ces questionnaires portent sur un large éventail de sujets : composition du ménage, santé de l'enfant et de la mère, habitat, famille, garde de l'enfant, alimentation de l'enfant, conditions de vie etc. Au terme de l'enquête auprès de la mère, l'enquêteur lui remettra soit un carnet de suivi alimentaire à compléter, soit un appareil de mesure du radon, soit un piège à poussières (sous-échantillons). Ces dispositifs devront être renvoyés par la mère au bout de 2 ou 3 mois à l'aide d'une enveloppe préaffranchie. Une des difficultés de cette première enquête tient dans le délai court entre la collecte des adresses en maternité et l'enquête ayant lieu 2 mois après la naissance. Le temps pour la saisie, la vérification, la transmission des adresses à l'Insee, l'établissement des contrats des enquêteurs par l'Insee en fonction des lieux d'habitation des mères est en effet extrêmement court.

Les vagues de collecte ultérieures

Il est prévu des enquêtes téléphoniques annuelles auprès de la mère et/ou du père. A certains âges clés du développement de l'enfant auront lieu des enquêtes en face-à-face de plus grande ampleur (à 3 ans par exemple) et des bilans médicaux (à 6 ans probablement). Un couplage à des données administratives est également prévu, notamment les données de l'assurance maladie.

3. Évaluation de la non-réponse : résultats des études pilotes

Les études pilotes débutées en 2007 dans 4 régions de France ont permis de tester la faisabilité et l'acceptation d'une telle étude. Les enfants inclus dans ces études pilotes seront suivis dans le temps afin de tester les étapes ultérieures de la collecte. Les résultats de participation présentés concernent les premières étapes de l'étude : en maternité et à 2 mois.

3.1 Ampleur de la non-réponse totale

Le premier type de non-réponse totale concerne la non-participation des maternités à l'enquête. Sur un total de 75 maternités sollicitées, 62 ont accepté de participer, soit 83 %. La participation des maternités a surtout été faible dans un département français (la Seine-Saint-Denis) avec un refus de 35 % des maternités qui peut s'expliquer par un contexte sanitaire et social plus difficile dans ce département.

Le second type de non-réponse concerne la non-réponse individuelle des mères à l'enquête dans les maternités participantes (tableau 3.1-1). On observe un taux de recueil d'informations du dossier médical de 95 % des naissances. Parmi ces mères pour lesquelles l'information du dossier médical a été recueillie par les sages-femmes, 88 % ont accepté de répondre à l'entretien en maternité (ce qui correspond à 85 % des naissances vivantes ciblées pour cet entretien). Ces taux élevés étaient attendus compte tenu du fait que ce type d'enquête est réalisé régulièrement (tous les 3-5 ans) dans les maternités (Enquêtes nationales périnatales). Ces informations sont

précieuses car elles nous fournissent des données pour une grande partie des non-répondants à l'enquête Elfe et permettent la comparaison des répondants et des non-répondants utile pour les ajustements pour la non-réponse.

L'acceptation d'Elfe, impliquant le suivi dans le temps, est plus faible, d'environ 55 % des naissances vivantes dans les maternités participantes. Ce taux est néanmoins satisfaisant compte tenu du contexte difficile d'une inclusion en maternité (surcharge de travail en maternité mais aussi fatigue de la mère après la naissance). Les prélèvements ont également été bien acceptés puisque respectivement 98 %, 91 % et 81 % des mères qui ont accepté le suivi Elfe ont accepté respectivement le recueil de ces prélèvements, la conservation de ces prélèvements pour des études ultérieures et l'utilisation de ces prélèvements pour d'éventuelles études génétiques. Pour des raisons pratiques, les prélèvements n'ont cependant pas pu être réalisés pour l'ensemble des mères, notamment pour le lait maternel (réduction du temps de séjour à la maternité avant la montée de lait).

Tableau 3.1-1
Résultats de participation en maternité.

	Mères					
	Pilote 1		Pilote 2		Total	
	nombre	%	nombre	%	nombre	%
Total des naissances	340	100	560	100	900	100
Recueil dossier médical	318	94	538	96	856	95
Naissances vivantes	338	100	552	100	890	100
Enquête en maternité	294	87	461	84	755	85
Acceptation suivi Elfe	191	57	301	55	492	55

Parmi les mères pour lesquelles on dispose du recueil d'informations dans le dossier médical, près de 50 % des non-réponses à l'entretien en maternité sont des refus et 22 % sont dues à des difficultés liées à la langue. Viennent ensuite une sortie de la mère avant le contact avec l'enquêtrice (12 %), des difficultés liées à l'état de santé de l'enfant (9 %) ou de la mère (3 %).

Parmi les mères ayant répondu à l'entretien en maternité et n'ayant pas accepté le suivi ultérieur pour Elfe, 83 % sont des refus, 7 % des difficultés liées de langue qui ont compliqué l'explication du projet et le consentement, et 10 % d'autres raisons (résultats du second pilote uniquement). A noter que 10 % de ces refus sont exprimés par le père, ce qui souligne l'importance de mieux impliquer et informer le père sur l'étude.

Le tableau 3.1-2 présente les résultats de participation à l'enquête à 2 mois. Ces résultats sont donnés sur la base des adresses exploitées c'est-à-dire que sont exclus les foyers hors des régions pilotes (ce qui ne sera pas le cas pour l'enquête nationale) et quelques adresses inexploitées. Parmi les familles ayant été sollicitées par l'Insee, 86 % des mères et 79 % des pères ont accepté de participer. On observe 7 % de refus pour le questionnaire auprès de la mère et 11 % de refus pour le questionnaire auprès du père, soit 54 % des non-réponses dans les deux cas ; 5 % des mères et 6 % des pères n'ont pas pu être joints.

Tableau 3.1-2
Résultats de participation à l'enquête aux 2 mois de l'enfant.

	Pilote 1		Pilote 2		Total	
	nombre	%	nombre	%	nombre	%
Acceptations en maternité	191	...	301	...	492	...
Adresses exploitées	171	100	277	100	448	100
Questionnaire mère						
Réponse	154	90,1	233	84,1	387	86,4
Refus	13	7,6	20	7,2	33	7,4
Impossible à joindre	4	2,3	17	6,1	21	4,7
Autre non-réponse	0	0,0	7	2,5	7	1,6
Questionnaire père						
Réponse	146	85,4	206	74,4	352	78,6
Refus	16	9,4	31	11,2	47	10,5
Impossible à joindre	7	4,1	21	7,6	28	6,3
Autre non-réponse	2	1,2	10	3,6	12	2,7
Motif de non-réponse inconnu	0	0,0	10	3,6	10	2,2

... n'ayant pas lieu de figurer

3.2 Facteurs liés à la non-réponse

Les taux d'acceptation d'Elfe en maternité varient très fortement en fonction des maternités avec un minimum de 4 % et un maximum de 100 % suggérant un effet fort de la maternité ou de l'enquêteur, les deux étant difficiles à distinguer, chaque enquêteur étant généralement affecté à une seule maternité et chaque maternité ayant en général un ou deux enquêteurs.

Par ailleurs, on observe pour le premier pilote une décroissance du taux d'acceptation avec le niveau d'équipement de la maternité². Les taux sont plus élevés dans les maternités de niveau 1 (89 % ; variant de 40 % à 100 % selon les maternités ; 15 maternités concernées) que dans celles de niveau 2 (57 % ; variant de 19 % à 100 % selon les maternités ; 14 maternités concernées) ou 3 (29 % ; variant de 17 % à 48 % selon les maternités ; mais seulement 3 maternités concernées). Le taux de réponse décroît par ailleurs linéairement lorsque le nombre de naissances dans la maternité augmente (ce facteur est très lié au niveau de la maternité). Ceci n'est observé que pour le premier pilote, ce qui pourrait s'expliquer au moins en partie par des situations plus contrastées dans les maternités pour le second pilote (département de la Seine-Saint-Denis).

Nous avons ensuite comparé le taux d'acceptation du suivi Elfe en maternité selon les caractéristiques sociodémographiques des mères (tableau 3.2-1). Cette première analyse exploratoire est réalisée sur l'ensemble des mères ayant répondu à l'enquête en maternité, ce qui explique que les taux présentés soient plus élevés que le taux global d'acceptation (taux de 55 % présenté en 3.1). La participation en maternité augmente avec l'âge puis décroît pour les plus de 35 ans ($p=0,04$), augmente avec le niveau d'études ($p=0,03$), est plus faible chez les ouvriers, employés, sans profession ($p<0,01$), est plus faible chez les mères de nationalité étrangère ($p<0,01$), augmente avec le nombre d'accouchements antérieurs et décroît après 3 accouchements antérieurs ($p=0,01$).

Ces résultats exploratoires sont utiles pour une meilleure compréhension des mécanismes de non-réponse et pour la planification des ajustements pour la non-réponse (cf. paragraphe 4.3).

² On définit 3 niveaux d'équipement des maternités : les grossesses sans complications sont orientées vers un niveau 1 ; les niveaux 2 comportent une unité de néonatalogie ; les niveaux 3 disposent d'un service de réanimation néonatale.

Tableau 3.2-1

Taux d'acceptation de l'étude Elfe en maternité en fonction des caractéristiques sociodémographiques des mères recueillies lors de l'entretien en maternité.

	Taux d'acceptation	Effectif	p (Chi-deux)
Age de la mère			
<25 ans	54 %	92	0,04
25 ans - 29 ans	61 %	234	
30-34 ans	70 %	262	
≥ 35 ans	65 %	157	
Niveau d'études			
Non-scolarisé/Primaire/Collège	58 %	78	0,03
Enseignement professionnel court	61 %	143	
Niveau bac	61 %	173	
Niveau bac +1 ou 2 ans	65 %	189	
Niveau bac + 3 ans et plus	75 %	155	
Catégorie socioprofessionnelle			
Sans profession	53 %	125	<0,01
Cadres, artisans, commerçants	76 %	84	
Professions intermédiaires	70 %	132	
Employés	64 %	316	
Ouvriers	63 %	82	
Nationalité			
Française	66 %	641	<0,01
Etrangère	51 %	104	
Nombre d'accouchements antérieurs			
0	60 %	291	0,01
1	64 %	261	
2	76 %	123	
3 et plus	57 %	60	

4. Pondérations et traitement de la non-réponse

4.1 Population cible et principe des pondérations

L'objectif principal du projet est de permettre la production d'estimations longitudinales. La population cible longitudinale est la même à chaque vague de collecte et correspond à la population cible à l'inclusion c'est-à-dire les naissances vivantes en France de l'année d'inclusion. Des pondérations longitudinales seront produites à chaque vague de collecte et pour chaque mode de collecte afin de permettre la production d'estimations représentatives de cette population.

De manière classique le processus d'établissement de ces pondérations se fait en 3 étapes : 1) Établissement des pondérations initiales 2) Ajustements pour la non-réponse 3) Calage sur marges.

La nécessité de produire des pondérations transversales (pour être représentatif de la population des enfants d'un âge donné en France à un moment donné, par exemple des enfants âgés de 5 ans en 2015) doit encore être évaluée et dépendra de l'information disponible et de l'inclusion ultérieure d'enfants non-ciblés en 2010 (enfants issus de l'immigration par exemple).

4.2 Pondérations initiales

Les pondérations initiales correspondent aux pondérations initiales de l'EDP et sont égales au nombre total de jours de l'année divisé par le nombre de jours sélectionnés pour l'enquête :

$$w_{initial} = \frac{365}{16}.$$

Une correction pourrait être appliquée pour tenir compte du fait que les naissances ne sont pas réparties uniformément sur l'année (fluctuations saisonnières et hebdomadaires).

Les naissances à domicile ne pourront pas être approchées sauf celles qui feront l'objet d'un transfert en maternité après la naissance. Ce défaut de couverture, qui devrait représenter moins de 1 % des naissances sera traité par calage sur les données d'état civil (cf. paragraphe 4.4).

4.3 Ajustement pour la non-réponse totale

La non-réponse totale sera traitée préférentiellement par des méthodes de repondération. Les non-réponses totales de pères (lorsque la mère répond) pourraient cependant être traitées par imputation (cf. paragraphe 4.5). La production de deux jeux de pondérations différents (pour le père et la mère) compliquerait en effet l'utilisation des pondérations dans les analyses statistiques, en particulier si l'on utilise simultanément des données du père et des données de la mère, et il serait préférable de n'avoir qu'un seul jeu de pondérations.

Le premier type de non-réponse à traiter est la non-réponse totale due à la non-participation des maternités. Les poids seront ajustés pour tenir compte de la non réponse des maternités en multipliant le poids initial par le nombre de naissances total dans les maternités pendant les 16 jours divisé par le nombre de naissances dans les maternités participantes pendant les 16 jours, ce qui nécessite de connaître le nombre de naissances dans les maternités participantes et non-participantes (ces dernières seront contactées pour connaître le nombre de naissances).

La non-réponse totale individuelle sera ensuite traitée à chaque vague de collecte. L'ajustement sera réalisé par constitution de Groupes de Réponse Homogènes (GRH) qui regroupent des individus dont la probabilité de réponse est supposée homogène. Les informations disponibles dans l'enquête réalisée en maternité et dans le dossier médical pour les répondants mais également une grande partie des non-répondants seront utilisées. Les analyses des facteurs liés à la non-réponse présentées en section 3 nous donnent une première idée des facteurs pouvant être utilisés pour réaliser cet ajustement : âge de la mère, catégorie socioprofessionnelle de la mère, nationalité, niveau d'études, nombre de naissances antérieures, caractéristiques de la maternité (nombre de naissances, niveau d'équipement) etc.

L'ajustement pour la non-réponse est alors réalisé au sein des GRH : le poids initial (ajusté pour la non-réponse des maternités) est divisé par le taux de réponse observé dans le GRH.

$$w_{ajustement\ non-réponse} = \frac{w_{initial}}{TR_{GRH}} \quad \text{où } TR_{GRH} \text{ est le taux de réponse observé au sein du GRH}$$

Des tests seront réalisés en 2009 pour déterminer la méthode de constitution des GRH la plus adaptée (segmentation, méthodes des scores...).

Un grand nombre d'information étant recueilli à 2 mois (questionnaires mais également parodonnées), il semble judicieux d'utiliser cette information pour réaliser les ajustements pour la non-réponse aux vagues de collecte ultérieures. Par rapport aux facteurs utilisés à la première vague, des informations telles que la composition du ménage, le revenu, des informations sur le logement (type de logement, locataire/propriétaire) pourront être utilisées. Des parodonnées récoltées lors de l'enquête à 2 mois pourront également être mobilisées (nombre de contacts, heure du dernier contact etc.) (Beaumont, 2005).

4.4 Calage sur marges

Des informations auxiliaires issues de sources externes pourront également être utilisées pour améliorer la précision des estimations et caler l'échantillon sur des totaux connus par méthode de calage sur marges (Deville et Särndal, 1992). Le choix et la disponibilité des sources externes est en cours d'évaluation et pourrait comprendre notamment les données de l'état civil qui fournissent des informations sur l'ensemble des naissances vivantes et sans vie en France (Insee, 2007). Ces données contiennent notamment des informations sur la date et le lieu de naissance, le sexe de l'enfant, des renseignements sur la mère et le père (notamment leur âge, profession et nationalité), les conditions d'accouchement (établissement spécialisé/à domicile ou autre lieu avec assistance/sans assistance) et le type d'accouchement (naissance simple ou multiple). En pratique, la macro SAS CALMAR2 (Sautory, 2003) sera utilisée.

4.5 Imputation

De manière classique, l'imputation est la méthode privilégiée pour le traitement des non-réponses partielles. Il semble cependant que dans certains cas, l'imputation soit également utile pour traiter la non-réponse totale notamment en présence de non-réponse de vague ou lorsque par exemple à une même vague, la mère répond et le père ne répond pas. Une imputation par donneur (le donneur est un individu répondant choisi en fonction de ses caractéristiques) peut sembler plus judicieuse dans ce cas à l'instar de ce qui a pu être fait dans certaines études longitudinales canadiennes (Simard, 2008).

Pour les non-réponses d'items au sein de chaque questionnaire, l'imputation systématique tout en préservant les relations entre les variables semble difficile du fait de l'ampleur du projet et de son caractère largement pluridisciplinaire. Des analyses plus précises de la non-réponse partielle dans les études pilotes seront réalisées en 2009 afin d'étudier l'opportunité de réaliser des imputations pour tout ou partie des variables.

4.6 Estimation de la variance

Dans le cadre d'enquêtes à plan de sondage complexe et en présence d'ajustements pour la non-réponse et d'imputation, les méthodes classiques d'estimation de la variance peuvent mener à des estimations erronées (sous-estimations des variances des estimations). Dans ce contexte, les méthodes bootstrap offrent des alternatives aux méthodes classiques faciles à mettre en œuvre puisqu'elles reposent sur une méthode de rééchantillonnage dans l'échantillon observé et sur une formule unique de calcul de la variance des estimations (Rao, Wu, Yue, 1992). Pour Elfe en particulier, qui se caractérise par un plan de sondage simple mais en présence d'ajustements pour la non-réponse, de calage et éventuellement d'imputation, le bootstrap avec remise semble être la méthode la plus adaptée. La difficulté résidera dans l'explication de la bonne utilisation des poids bootstrap pour un public très large d'utilisateurs, de culture et de niveau statistique varié.

5. Conclusion

Les études pilotes nous fournissent de premières estimations de la non-réponse attendue et ces premiers résultats sont satisfaisants pour des études pilotes et dans le contexte d'une inclusion en maternité. Ces études pilotes nous ont permis d'identifier des possibilités d'amélioration des taux de réponse (formation des enquêtrices en maternité, information des différents acteurs en maternité, manque d'information et d'implication du père source de refus etc.) mais a aussi souligné les difficultés des réalités du terrain (manque de personnel etc.).

La non-réponse est très variable en fonction de la maternité ce qui peut traduire un effet enquêteur, mais semble également révéler des difficultés plus ou moins importantes à réaliser l'enquête dans certaines maternités (surcharge de travail, communication entre les équipes). On observe également une sélection de l'échantillon selon certaines caractéristiques sociodémographiques des mères : âge, niveau d'études, profession, nationalité, nombre d'accouchements antérieurs. Ces premiers résultats sont utiles pour comprendre les mécanismes de non-réponse et envisager les stratégies de traitement des non-réponses pour l'enquête en vraie grandeur. Les perspectives sont de poursuivre les analyses de la non-réponse notamment par une modélisation multivariée intégrant les différents

facteurs liés à la non-réponse. La méthodologie de traitement des non-réponses va également être affinée et finalisée en 2009-2010.

Remerciements

L'enquête Elfe est une réalisation conjointe de l'Institut national d'études démographiques (Ined), de l'Institut de veille sanitaire (InVS), de l'Institut national de la santé et de la recherche médicale (Inserm), de l'Institut national de la statistique et des études économiques (Insee), de la Direction générale de la santé (DGS, Ministère en charge de la santé), de la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees, Ministère en charge de la santé et de l'emploi), de la Direction de l'évaluation, de la prospective et de la performance (Depp, Ministère en charge de l'Enseignement) et de la Caisse nationale des allocations familiales (Cnaf), avec le soutien du Ministère de la recherche, de l'Agence nationale de la recherche (ANR), du Comité de concertation pour les données en sciences humaines et sociales (CCDSHS), de la Mutuelle générale de l'Education nationale (MGEN), du Ministère de la Culture (Deps).

Bibliographie

- Beaumont, J.F. (2005). L'utilisation de renseignements sur le processus de collecte de données pour traiter la non-réponse totale au moyen de l'ajustement de poids, *Techniques d'enquête*, 31, 249-254.
- De Leeuw, E. et De Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison, *Survey Nonresponse* (Éds. R.M. Groves et coll.), Wiley: New York, 41-54.
- Deville, J.C. et Särndal, C.E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87, 376-382.
- Insee (2007), Les sources statistiques de l'État civil, *Sources et méthodes*.
- Rao, J.N.K., Wu, C.F.J. et Yue, K. (1992). Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes, *Techniques d'enquête*, 18, 225-234.
- Richet-Mastain, L. (2007), Bilan démographique 2006 : un excédent naturel record, *Insee Première*, 1118.
- Sautory, O. (2003). Calmar 2 : Une nouvelle version du programme Calmar de redressement d'échantillon par calage, *Recueil : Symposium 2003, Défis reliés à la réalisation d'enquêtes pour la prochaine décennie*, Statistique Canada.
- Simard, M. (2008). L'imputation dans les enquêtes longitudinales au Canada, *Méthodes de sondages* (Éds. Guibert et coll.), Dunod, 45-52.