

**Méthodologie statistique**

**M 2015/01**

**La collecte multimode et le paradigme  
de l'erreur d'enquête totale**

**Tiaray Razafindranovona**

**Document de travail**



Institut National de la Statistique et des Études Économiques



**M 2015/01**

**La collecte multimode et le paradigme  
de l'erreur d'enquête totale**

**Tiaray Razafindranovona \***

Ce document a bénéficié des commentaires, corrections et remarques de nombreuses personnes. Je les en remercie et tout particulièrement Gaël de Peretti pour son aide et son soutien tout au long de ce travail, ainsi que François Beck, Stéphane Gregoir, Stéphane Legleye et Olivier Sautory. Je reste seul responsable des erreurs qui pourraient y demeurer.

---

\*DMCSI

18, bd Adolphe Pinard - 75675 PARIS CEDEX 14

Direction de la méthodologie et de la coordination statistique et internationale -Département des Méthodes Statistiques - Timbre L101 - 18, bd Adolphe Pinard - 75675 PARIS CEDEX - France -  
Tél. : 33 (1) 41 17 66 01 - Fax : 33 (1) 41 17 66 33 - CEDEX - E-mail : [L001-dg@insee.fr](mailto:L001-dg@insee.fr) - Site Web Insee : <http://www.insee.fr>

*Ces documents de travail ne reflètent pas la position de l'Insee et n'engagent que leurs auteurs.  
Working papers do not reflect the position of INSEE but only their author's views.*

# La collecte multimode et le paradigme de l'erreur d'enquête totale

Tiaray Razafindranovona \*

## Résumé

Les enquêtes multimode ne sont pas une nouveauté. Depuis longtemps, sont associés au sein d'une même enquête, des entretiens en face à face et des entretiens téléphoniques, voire des questionnaires auto-administrés. Cependant, depuis quelques années, se développe le recours à internet. D'abord utilisé auprès des entreprises, ce mode de collecte commence à se diffuser dans les enquêtes auprès des ménages, en particulier du fait de son avantage en termes de coût. Toutefois, avant de généraliser l'utilisation d'internet dans les enquêtes ménages, en tant que mode de collecte privilégié ou complémentaire, il est important de bien cerner les questions méthodologiques sous-jacentes, dans un cadre général d'analyse bien adapté, qui est celui de l'erreur d'enquête totale. En particulier, la question des effets de mode est cruciale dès lors que l'on envisage l'agrégation, dans le cadre d'une collecte multimode, de réponses provenant de modes différents.

**Mots-clés :** enquête multimode, collecte par internet, effets de mode, méthodologie d'enquête, erreur d'enquête totale.

## Abstract

Mixed-mode surveys are not that new. For a long time, face to face and telephone interviews or even auto-administered questionnaires have been combined in the same survey. Yet, in recent years, web data collection has expanded. First used for business surveys, this mode of data collection begins to spread into households surveys mainly for budgetary reasons. But before generalizing the use of internet in households surveys, as a privileged or complementary mode of data collection, underlying methodological questions have to be clearly defined. And the paradigm of the total error survey is the theoretical framework that such an investigation suits. In particular, in the case of a mixed-mode survey, the question of mode effects seems crucial if aggregating responses from different modes of data collection is to be considered.

**Keywords :** mixed-mode survey, web data collection, mode effects, survey methodology, total survey error.

---

\* INSEE, Département des méthodes statistiques, tiaray.razafindranovona@insee.fr

# Table des matières

<b>Introduction</b>	<b>4</b>
<b>1 L'erreur d'enquête totale</b>	<b>5</b>
1.1 Le cumul des différentes erreurs d'enquête . . . . .	6
1.2 Différentes typologies . . . . .	6
1.3 Les principales composantes de la TSE . . . . .	7
1.3.1 L'erreur de spécification . . . . .	7
1.3.2 L'erreur de couverture . . . . .	8
1.3.3 L'erreur d'échantillonnage . . . . .	9
1.3.4 L'erreur de non-réponse . . . . .	9
1.3.5 L'erreur de mesure, d'observation . . . . .	11
1.3.6 L'erreur de saisie, de codification . . . . .	14
1.4 Éléments de contexte historique . . . . .	15
1.5 Les apports du concept . . . . .	16
1.6 Limites et extensions . . . . .	17
1.6.1 Précision ou qualité ? . . . . .	17
1.6.2 La qualité totale d'enquête . . . . .	17
1.7 Panorama des questions de recherche . . . . .	18
1.7.1 Les effets enquêteurs . . . . .	18
1.7.2 Protocoles d'enquêtes adaptatifs . . . . .	19
1.7.3 Comparaison des différents modes de collecte . . . . .	19
1.7.4 Arbitrages erreur de couverture, erreur de non-réponse . . . . .	19
1.7.5 Interactions entre erreurs de non-réponse et erreurs de mesure . . . . .	20
<b>2 La collecte multimode à travers le prisme de l'erreur d'enquête totale</b>	<b>21</b>
2.1 Un peu de vocabulaire . . . . .	21
2.2 Pourquoi le multimode ? . . . . .	23
2.3 Multimode et TSE . . . . .	23
2.3.1 TSE : un cadre adapté ? . . . . .	23
2.3.2 Composantes d'erreur et solutions multimodes . . . . .	24
2.4 Comparaison entre les modes . . . . .	24
2.4.1 Couverture . . . . .	24
2.4.2 Taux de réponse . . . . .	26
2.4.3 Erreur de mesure . . . . .	26
2.4.4 Coûts . . . . .	27
2.4.5 La collecte par internet . . . . .	27

2.4.6	Une revue de littérature . . . . .	29
<b>3</b>	<b>L'ampleur des effets de mode</b>	<b>32</b>
3.1	Comparaison naïve . . . . .	32
3.2	Contrôle par les observables . . . . .	33
3.2.1	Exemples de contrôles par les observables . . . . .	33
3.2.2	Des effets toujours confondus ? . . . . .	34
3.3	Expérimentation contrôlée . . . . .	35
3.4	Autres exemples de protocoles . . . . .	36
3.5	Intérêt des expérimentations . . . . .	37
3.6	Corrections, ajustements . . . . .	38
3.7	Quelques résultats sur les effets de mode . . . . .	38
3.7.1	Des répondants sur internet plus pessimistes . . . . .	39
3.7.2	Internet et papier : des différences ? . . . . .	39
	<b>Conclusion</b>	<b>41</b>
	<b>Bibliographie</b>	<b>42</b>

# Introduction

Les instituts en charge d'enquêter les ménages font aujourd'hui face à de nombreux défis. Ils mettent en oeuvre des protocoles toujours plus innovants pour répondre à la demande sociale et aux exigences déontologiques : respect de la confidentialité, interrogation sur des thèmes très sensibles ou encore études sur des populations minoritaires (Verger, 2013). L'exigence de qualité (Desrosières, 2003), dans toutes ses dimensions, est également très forte : précision, pertinence, cohérence, comparabilité ou encore fraîcheur (de Peretti et Razafindranovona, 2013). Mais alors que la demande sociale est toujours plus exigeante, les ménages sont de plus en plus difficiles à atteindre. Parmi les raisons que l'on peut évoquer, on peut citer les nombreuses barrières technologiques ou physiques (répondeur, affichage du numéro de téléphone entrant, interphones, digicodes) qui existent entre l'organisme en charge de l'enquête et le ménage à interroger qui peut filtrer les tentatives de contact, ou encore les effets de lassitude d'enquêtés de plus en plus sollicités.

Pour faire face à ces défis, les instituts engagent des réflexions pour adapter leurs processus de production et leurs protocoles de collecte. Ainsi, la coopération des enquêtés pourrait être recherchée par différents moyens de communication, en élargissant l'éventail des modes de réponse, par exemple en proposant la possibilité de répondre par internet.

Pour les instituts, ces réflexions sur la refonte de leurs processus de production des enquêtes ménages se déroulent dans un contexte particulièrement difficile du point de vue budgétaire. On ne peut occulter le fait que derrière la volonté d'aller vers des formes multimodes de collecte, il existe également des logiques de rationalisation des coûts (de Peretti et Razafindranovona, 2014).

Mais, au-delà des questions de coût, il est essentiel de bien cerner les questions méthodologiques sous-jacentes de précision et de qualité des enquêtes et de définir le cadre le plus adapté pour répondre à ces questions. On montrera ainsi que le paradigme de l'erreur d'enquête totale est le cadre qui convient pour analyser au préalable la pertinence ou non de ces différentes évolutions des processus de production.

# Chapitre 1

## L'erreur d'enquête totale

Le concept d'erreur d'enquête totale (TSE pour l'anglais *Total Survey Error*) est l'un des paradigmes dominants en méthodologie d'enquête. Il s'agit d'un cadre théorique utilisé pour optimiser les enquêtes statistiques : la grandeur cumulée des erreurs provenant de différentes sources au cours du processus d'enquête (conception, plan de sondage, collecte, contrôles, diffusion) doit être minimisée étant donné certaines contraintes (Lyberg, 2012), par exemple, de budget ou de temps. En effet, la théorie montre que si les enquêtes par sondage peuvent fournir des estimateurs sans biais des paramètres d'intérêt, de nombreuses possibilités de déviation des estimateurs par rapport à la "vraie" valeur sous-jacente existent. L'erreur d'enquête totale s'exprime traditionnellement sous la forme d'une erreur quadratique moyenne, somme des erreurs systématiques (biais) au carré et des erreurs variables (variance).

En pratique, la recherche de l'enquête optimale via un critère d'erreur totale n'est pas aisée. En effet, certaines composantes d'erreur d'enquête ne peuvent être mesurées directement, en particulier celles autres que l'erreur d'échantillonnage ; ainsi des investigations méthodologiques lourdes vont parfois être nécessaires pour connaître l'ampleur de ces erreurs. En particulier, distinguer les biais liés aux différentes composantes de l'erreur d'enquête totale n'est pas évident car les biais estimés sont conditionnels alors que les composantes de l'erreur d'enquête ne sont pas indépendantes : la décomposition de l'erreur d'enquête totale constitue un objet idéal qu'il est difficile d'atteindre. Les expérimentations réalisées pour accumuler de la connaissance dans le domaine ont été nombreuses ces dernières années et si les résultats ne sont pas toujours généralisables, ils peuvent être mobilisés pour donner une idée du sens et de l'ampleur possible des biais en jeu.

Ainsi, même si une mesure fine des différentes composantes n'est pas toujours possible, se placer dans le cadre de l'erreur d'enquête totale permet de bien définir les sources d'erreur potentielles et force à se poser la question de leur importance relative. Cela devrait constituer un préalable dès lors que l'on souhaite concevoir une enquête ne serait-ce que pour avoir une idée des conséquences sur la validité des estimations. Le cadre de l'erreur d'enquête totale est également bien adapté dès lors que l'on souhaite modifier un protocole ou encore comparer différents protocoles d'enquête.



## 1.1 Le cumul des différentes erreurs d'enquête

Les données d'enquête sont sujettes à des erreurs pouvant provenir de différentes sources. L'erreur d'enquête totale constitue l'accumulation de toutes ces erreurs d'enquête. Ces dernières peuvent se situer lors de la phase de conception, au niveau du plan de sondage, à la collecte ou encore lors des étapes postérieures à cette dernière (contrôles, saisie, codification, diffusion).

Ainsi, il existe un certain consensus sur les principales composantes de l'erreur d'enquête totale, qui sont :

- l'erreur de couverture ;
- l'erreur d'échantillonnage ;
- l'erreur de non-réponse ;
- l'erreur de mesure ou d'observation.

D'autres composantes sont moins souvent étudiées dans la littérature, mais peuvent être intégrées dans l'erreur d'enquête totale : en amont, les erreurs de spécification, en aval les erreurs de saisie, de codification. Certains auteurs préconisent de tenir compte également des erreurs faites lors de la phase d'exploitation, par exemple lorsque les données d'enquêtes ont pour *output* immédiat des tableaux de résultats.

Nous verrons plus en détail les définitions ainsi que des illustrations de ces différentes composantes d'erreur.

## 1.2 Différentes typologies

Différentes typologies de l'erreur d'enquête totale existent dans la littérature.

Une première typologie oppose les erreurs d'échantillonnage aux autres types d'erreurs (voir figure 1.1). Les erreurs d'échantillonnage relèvent du principe même de l'enquête par sondage : comme seule une partie de la population cible est interrogée, une erreur est commise, qui diminue lorsque la taille de l'échantillon augmente. Les erreurs autres que l'échantillonnage peuvent en revanche augmenter avec la taille de l'échantillon.

Une autre typologie oppose les erreurs d'observation aux autres types d'erreurs. Dans cette typologie, on a d'un côté les erreurs liées à la mesure de la variable d'intérêt de l'enquête, de l'autre les erreurs qui relèvent plutôt des phénomènes d'exclusion d'une partie de la population cible dans le processus d'enquête (non couverts, non sélectionnés, non répondants).

Enfin, les erreurs peuvent être distinguées selon qu'elles sont plutôt systématiques (biais) ou variables (variance). Les erreurs d'échantillonnage constituent l'essentiel des erreurs variables alors que les erreurs systématiques proviennent plutôt des erreurs autres que l'échantillonnage. Mais des erreurs systématiques peuvent également relever de l'échantillonnage ; c'est le cas par exemple lorsque les calculs de précision reposent sur des formules ne correspondant pas au plan de sondage mis en oeuvre. Les erreurs autres que l'échantillonnage peuvent également contribuer à l'erreur variable, par exemple l'erreur de mesure qui relève de l'effet enquêteur ou

l'indifférence des répondants à une question, qui génère des réponses aléatoires.

L'erreur d'enquête totale s'exprime traditionnellement sous la forme d'une erreur quadratique moyenne; il s'agit de l'espérance au carré de la différence entre l'estimateur  $\hat{\theta}$  et la vraie valeur du paramètre d'intérêt  $\theta$ .

$$TSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = \text{Biais}^2(\hat{\theta}) + \text{Variance}(\hat{\theta})$$

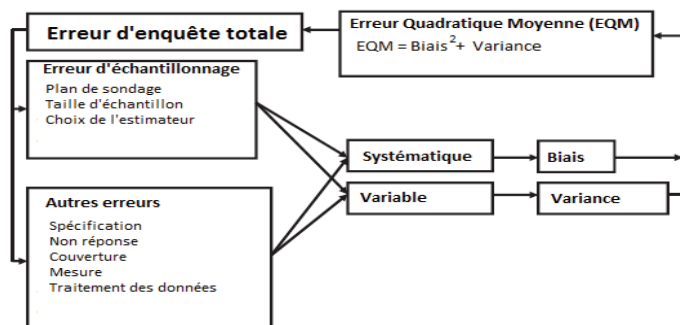


FIGURE 1.1 – Exemple de typologie (Biemer, 2010)

## 1.3 Les principales composantes de la TSE

### 1.3.1 L'erreur de spécification

Il s'agit de l'erreur commise quand la variable mesurée par l'enquête diffère de la variable d'intérêt. Ainsi, quand la variable construite ne correspond pas au phénomène que l'on souhaite mesurer, l'inférence et l'interprétation qui s'en suit risquent de ne plus être valides. Ce type d'erreur est lié aux concepts et aux objectifs en jeu dans l'enquête et peut résulter d'une mauvaise communication entre utilisateurs et concepteurs d'enquête.

#### Illustration d'un problème de spécification

Dans la plupart des cas, les évolutions importantes des questionnaires des enquêtes récurrentes relèvent d'adaptations à des évolutions de définitions des concepts sous-jacents. Ainsi, l'enquête Emploi de l'Insee s'est adaptée, au cours de son histoire, à des changements de définition et de mesure du chômage induits par les recommandations du Bureau International du Travail (Goux, 2003).

Mais il existe également des modifications de questionnaires qui répondent plutôt à des problèmes de spécification. Ainsi, en 1994, une question de l'enquête *Current Population Survey*<sup>1</sup> impliquant le terme "*layoff*" a été modifiée car ce terme pouvait être compris par les enquêtés

1. L'enquête Current Population Survey (CPS) est une enquête américaine mensuelle menée auprès d'environ 60 000 ménages que l'on peut rapprocher des *Labour Force Surveys* européens (dont la "déclinaison" française est l'enquête Emploi). Cette enquête est une source primaire de données sur les caractéristiques de la population active aux États-Unis.

comme licenciement permanent alors que le concept que les économistes souhaitaient mesurer était plutôt celui de la perte temporaire d'emploi (*layoff and recall*) (Biemer, 2010).

### 1.3.2 L'erreur de couverture

L'erreur de couverture est liée au fait que la base de sondage ne corresponde pas à la population cible. On parle de sous-dénombrement quand certaines unités de la population d'intérêt ont une probabilité nulle d'être sélectionnées.

Un problème de couverture se pose dès lors que certains éléments de la population cible ne peuvent être atteints par l'enquête ou ne peuvent répondre par les canaux proposés par l'organisation enquêtrice. La population cible dont il est question ici est celle sur laquelle on souhaite inférer des résultats à partir de l'enquête ; quand des unités de la population sont délibérément exclues pour des raisons qui tiennent aux objectifs de l'enquête ou pour des difficultés pratiques, il ne s'agit pas d'un problème de couverture en tant que tel.

L'expression du biais de couverture fait intervenir deux composantes : la proportion de la population non couverte, et la différence sur la variable d'intérêt entre population non couverte et population couverte.

$$B_{NC} = \frac{N_{NC}}{N} (\bar{Y}_C - \bar{Y}_{NC})$$

avec  $Y$  variable d'intérêt,  $N_{NC}$  population non-couverte,  $N$  population totale,  $\bar{Y}_{NC}$  moyenne de  $Y$  sur la population non-couverte et  $\bar{Y}_C$  moyenne de  $Y$  sur la population couverte.

#### Illustration d'un problème de couverture : l'enquête TIC

À l'Insee, l'enquête sur les Technologies de l'Information et de la Communication (TIC) auprès des ménages a vu son protocole évoluer régulièrement ces dernières années, principalement pour tenir compte des problèmes de couverture. En 2007, un échantillon unique de 7 000 ménages, tiré par Pages Jaunes Marketing Services (PJMS) dans la base des abonnés France Telecom, est interrogé par téléphone. La population cible n'est alors pas totalement couverte : il manque les ménages sur "liste rouge", les ménages uniquement équipés en téléphonie mobile ou encore les abonnés aux autres opérateurs de télécommunications. En 2008, pour mieux couvrir la population, outre l'échantillon tiré par PJMS dit principal, un échantillon complémentaire de ménages est tiré dans le recensement 2006 et les ménages qui ne se trouvent pas dans la base d'abonnés France Telecom sont interrogés en face à face.

En 2010, outre l'échantillon tiré par PJMS, un échantillon complémentaire de ménages est tiré dans les bases fiscales. Ces ménages sont contactés par courrier et peuvent répondre par papier ou par internet. Enfin, à partir de 2011, l'Insee utilise un logiciel (Annucapt) qui recherche automatiquement des numéros de téléphone sur le site internet des Pages blanches. Un échantillon de 16 000 ménages est tiré dans les bases fiscales et les ménages retrouvés dans l'annuaire sont interrogés par téléphone, tandis que ceux qui ne le sont pas sont contactés par courrier et peuvent répondre par papier ou sur internet. Par ailleurs, un échantillon de contrôle est également tiré : des ménages retrouvés dans l'annuaire sont interrogés par internet ou papier (Gombault et Duée, 2012).

### 1.3.3 L'erreur d'échantillonnage

Cette erreur est liée au fait que les estimations reposent sur les individus sélectionnés dans l'échantillon et non sur la population dans son ensemble.

Dans le cadre d'un sondage aléatoire simple (sans remise), l'erreur d'échantillonnage sur la moyenne de la variable d'intérêt  $Y$  estimée par la moyenne empirique d'échantillon  $\bar{y}$  est estimée par :

$$\widehat{Var}(\bar{y}) = (1 - f) \frac{s_{corr}^2}{n}$$

où

$$f = \frac{n}{N} \quad s_{corr}^2 = \frac{1}{n-1} \sum_{i \in S} (y_i - \bar{y})^2$$

avec  $n$  la taille de l'échantillon sélectionné,  $N$  la taille de la population cible et  $S$  l'ensemble des unités sélectionnées.

Cette composante de l'erreur d'enquête totale a été très largement étudiée dans la communauté des statisticiens et des méthodologues, on peut citer par exemple, comme ouvrages de références *Model Assisted Survey Sampling* (Särndal *et al.*, 2003) ou *Les techniques de sondage* (Ardilly, 2006).

### 1.3.4 L'erreur de non-réponse

Cette erreur est liée à l'absence de réponse de certains individus sélectionnés dans l'échantillon. La non-réponse est dite totale quand aucune information n'est collectée sur l'unité sélectionnée, partielle quand elle porte sur certaines variables. Souvent, la non-réponse se concentre parmi certains sous-groupes de la population et l'erreur de non-réponse peut être conséquente quand les variables d'intérêt de l'enquête sont également liées aux sous-groupes en question. La non-réponse est ainsi susceptible de détruire la représentativité des échantillons, car les non-répondants n'ont en général pas les mêmes caractéristiques que les répondants (Sautory, 2013).

Dans une première version dite déterministe, la réponse peut être considérée comme un *outcome* fixe et l'expression du biais de non-réponse fait intervenir deux composantes : la proportion de non-répondants et la différence sur la variable d'intérêt entre répondants et non-répondants.

$$B_{NR} = \frac{N_{NR}}{N} (\bar{Y}_R - \bar{Y}_{NR})$$

avec  $Y$  variable d'intérêt,  $N_{NR}$  population non-répondante,  $N$  population totale,  $\bar{Y}_{NR}$  moyenne de  $Y$  sur la population non-répondante et  $\bar{Y}_R$  moyenne de  $Y$  sur la population répondante.

Une autre manière d'exprimer ce biais est de considérer la réponse comme stochastique ; le biais va alors dépendre de la corrélation entre la variable d'intérêt et la probabilité de réponse,

ainsi que de la propension moyenne à répondre au sein de la population (Bethlehem, 1988).

$$B_{NR} \approx \frac{Cov(\rho, Y)}{\bar{\rho}}$$

avec  $Y$  variable d'intérêt,  $\rho$  propension aléatoire à répondre.

### Formes et facteurs de non-réponse

La non-réponse totale peut prendre différentes formes, on peut citer parmi celles-ci (Lynn, 2008) :

- l'impossibilité pour l'enquêteur de localiser ou d'identifier l'unité sélectionnée ;
- l'impossibilité pour l'enquêteur de contacter l'unité sélectionnée ;
- l'incapacité pour l'enquêté de répondre à l'enquête (problèmes de santé, absence) ;
- le refus de répondre à l'enquête ;
- l'incapacité des deux parties à communiquer (problèmes de langue) ;
- la perte de questionnaires, de données.

Notons que dans certains de ces cas, l'unité pour laquelle aucune information n'est collectée peut être considérée comme "Hors champ" par l'organisation enquêtrice.

Dans ce qui relève du refus "volontaire" de répondre à l'enquête, de nombreux facteurs entrent en jeu : le type d'enquête (obligatoire ou volontaire), le mode de collecte, la durée du questionnaire, la qualité du contact ou encore le thème de l'enquête. La non-réponse partielle est souvent liée à la nature de l'effort cognitif demandé : l'individu enquêté peut choisir de ne pas répondre à une question lorsqu'il ne parvient pas à la comprendre ou quand il juge que les efforts demandés pour y répondre sont trop importants (exemple des questions qui font appel à la mémoire). Les questions délicates ou sans rapport avec le thème du module de questions où elles s'insèrent (voire avec le thème principal de l'enquête), que l'enquêté peut juger comme indiscretes, peuvent également engendrer une non-réponse conséquente.

### Mécanismes de non-réponse

Dans l'optique de corriger au mieux le biais de non-réponse, une première étape est d'avoir une idée sur le type de modélisation de la non-réponse qui semble raisonnable. On distingue généralement trois types de mécanismes de non-réponse (Rubin, 1976; Little et Rubin, 2002) :

- uniforme ;
- ignorable ;
- non-ignorable.

Le mécanisme est dit uniforme (MCAR pour *Missing Completely At Random*) quand tous les individus ont la même probabilité de réponse. Il s'agit d'une hypothèse très forte, généralement peu réaliste, que l'on peut néanmoins émettre dans les cas où la perte d'information est "aléatoire" comme lorsque des questionnaires sont perdus.

On parle de mécanisme de réponse ignorable (MAR pour *Missing At Random*) quand les probabilités de réponse peuvent être expliquées à l'aide de variables auxiliaires disponibles.

$$P(r|y_o, y_m) = P(r|y_o)$$

avec  $r$  l'indicatrice de réponse,  $y_o$  les variables observées,  $y_m$  les variables manquantes. Cela signifie qu'étant donné l'information observée, le mécanisme ne dépend pas de l'information manquante.

Enfin, le mécanisme est dit non-ignorable (NMAR pour Non Missing At Random) lorsque la non-réponse va dépendre de la variable d'intérêt même une fois que l'on a pris en compte l'information auxiliaire.

À titre d'exemple et pour donner quelques ordres de grandeur, les taux de réponse d'enquêtes récurrentes de l'Insee comme l'enquête Emploi, l'enquête statistique sur les ressources et conditions de vie (dispositif SRCV) ou encore l'enquête de conjoncture auprès des ménages (Camme) sont d'environ 80 % en 2014.

### 1.3.5 L'erreur de mesure, d'observation

Une erreur de mesure ou d'observation est commise lorsque l'information collectée sur l'individu diffère de la "vraie" valeur sous-jacente de la variable. Les sources d'erreur potentielles sont diverses.

Par exemple, les enquêtés peuvent volontairement ou non fournir des réponses incorrectes aux questions posées. Ainsi, l'enquêté peut fournir une réponse incorrecte parce qu'il ne veut pas complètement se dévoiler. La réponse peut être involontairement incorrecte quand l'enquêté ne parvient pas à se souvenir de l'évènement sur lequel il est interrogé : c'est le problème des biais de mémoire.

Les enquêteurs peuvent influencer les réponses que vont donner les personnes interrogées par le biais de suggestions implicites ou de mécanismes psychologiques liés à l'interaction entre les deux parties. Les enquêteurs peuvent également sur-interpréter les réponses de l'enquêté à cause d'une perception sélective fonction de leurs propres opinions.

En outre, le questionnaire peut être également source d'ambiguïté ; dans la mesure du possible, les concepts utilisés doivent satisfaire les objectifs de l'enquête et permettre d'obtenir des réponses claires et cohérentes.

Deux thèmes sont souvent étudiés dans la littérature quand il est question de la qualité des réponses et du mode de collecte : la désirabilité sociale et la tendance au *satisficing* (faible implication du répondant), phénomènes qui font appel à des notions de psychologie du répondant.

#### La désirabilité sociale

La désirabilité sociale est la tendance des enquêtés à fournir des réponses qui donnent une bonne image d'eux-mêmes ou dont ils croient qu'elles se situent dans la norme sociale (Edwards, 1953). Ce mécanisme psychologique peut s'exercer de façon implicite, sans que l'on en ait conscience, ou au contraire être le résultat d'une volonté consciente de manipuler son image aux yeux des autres. Cela pose un problème en termes de validité des questionnaires collectés

dont les psychométriciens sont bien conscients depuis la fin des années 1930 (Paulhus, 1991). Les déterminants des biais de désirabilité sociale sont de plusieurs natures :

- thème de l'enquête : les biais sont susceptibles d'être importants lorsque la thématique d'une enquête est sensible (pratiques à risque, opinions politiques, etc.).
- questionnement : les questions et les modalités de réponse proposées doivent, dans la mesure du possible, respecter une certaine neutralité pour ne pas induire automatiquement une réponse qui se situerait dans la norme sociale.
- conditions d'anonymat et distance sociale avec l'enquêteur : les biais de ce type sont théoriquement décroissants avec la distance sociale (vis-à-vis de l'enquêteur) et sont également moins importants lorsque les conditions d'anonymat sont garanties. Par exemple, la réponse sera moins socialement dans la norme avec un enquêteur complètement anonyme qu'avec une personne présentant une certaine proximité sociale (par exemple un enquêteur habitant le même quartier). En particulier, ces biais vont dépendre de la manière dont le questionnaire est administré et donc du mode de collecte : l'ampleur des biais est *a priori* moins importante pour les questionnaires auto-administrés que pour les questionnaires réalisés par un enquêteur (de Leeuw et Dillman, 2008).

Notons également que plusieurs méthodes existent pour essayer de mesurer le biais de désirabilité sociale dont la plus connue est probablement l'échelle de Marlowe-Crowne (Crowne et Marlowe, 1960).

Le thème des pratiques sexuelles est une illustration classique du phénomène de désirabilité sociale lors d'une collecte d'information par questionnaire. En effet, sur ce thème sensible, les réponses ont des chances de se situer dans la gamme des comportements jugés acceptables d'après les normes sociales en vigueur (Langis et Germain, 2009) ou au contraire de générer une surdéclaration, chez les jeunes hommes en particulier (Lagrange, 1991).

Par exemple, Alexander et Fisher (2003) ont regardé quelles étaient les différences entre hommes et femmes dans les comportements sexuels déclarés, selon le mode de collecte de l'information. Pour cela, ils mettent en place trois protocoles différents, où le degré d'exposition du répondant diffère : dans un premier protocole, les participants répondent à un questionnaire en étant connectés à ce qu'on leur présente comme un détecteur de mensonge (protocole du *bogus pipeline*) ; dans le deuxième protocole, on leur dit que le détecteur de mensonge n'est pas actif pour les questions les plus sensibles sur la sexualité ; enfin dans le troisième protocole, le détecteur de mensonge n'est pas utilisé et les conditions d'anonymat ne sont pas assurées contrairement aux deux autres protocoles. Les différences entre hommes et femmes dans les comportements sexuels déclarés sont les plus importantes pour le troisième protocole qui est celui qui encourage le plus l'enquêté à se référer à des normes sociales quand il répond. Et les différences sont d'autant plus importantes que ces normes sociales sont genrées : par exemple, lorsque des pratiques sont socialement plus acceptables pour les hommes que pour les femmes (masturbation, exposition à du matériel audiovisuel de nature érotique ou pornographique). Ce sont surtout les déclarations des femmes qui sont sensibles au mode de collecte : par exemple, elles ont tendance à déclarer un faible nombre de partenaires sexuels dans le cadre du troisième protocole, bien moins important que celui déclaré par les hommes, alors que dans le cadre du premier protocole (*bogus pipeline*), il est légèrement plus élevé pour les femmes.

Un autre exemple de sous-déclaration chez les femmes a été constaté par André Béjin (1993) : l'exploitation de différentes questions sur la masturbation posées dans l'enquête Analyse des

comportements sexuels en France montre qu'il existe bel et bien certaines réticences à répondre à ce type de questions. En effet, l'examen des réponses à l'une des questions, non filtrée, alors ce que ce n'était pas prévu *a priori*, lui permet de mettre en évidence l'existence d'une sous-déclaration de la pratique de la masturbation.

### La faible implication du répondant (*satisficing*)

Le concept de *satisficing* est dû à l'économiste Simon (1956) qui a développé cette théorie ainsi que celle de la rationalité limitée qui sont en partie liées. Une des premières définitions qu'il donne est que le *satisficing* renvoie au fait d'arrêter de chercher le choix optimal dès lors que l'on a trouvé une option suffisante pour atteindre un certain niveau d'aspiration. Et plus précisément (Simon, 1982), les individus s'engagent dans le *satisficing* en choisissant une option qui permet d'atteindre un certain niveau spécifié sans que cette option ne soit forcément unique ou la meilleure. Ainsi le critère de *satisficing* s'oppose au critère de maximisation des modèles de choix classiques ou peut, au contraire, également être considéré comme une généralisation de ce critère.

Ainsi, dans un modèle classique, lorsque la décision repose sur un critère de maximisation, l'alternative  $x$  choisie est telle que  $f(x) \geq \text{Max } f[A]$  et donc que l'utilité pour cette alternative ne peut être inférieure au maximum atteignable en parcourant l'ensemble des alternatives possibles de la liste  $A$ . En revanche, avec le critère de *satisficing*, l'alternative  $x$  est choisie telle que  $f(x) \geq \theta(A)$  où  $\theta(A)$  est une valeur seuil inférieure ou égale au maximum atteignable sur la liste  $A$ .

Le *satisficing* prend un sens particulier lorsqu'on utilise ce terme dans le cadre des enquêtes statistiques. Le *satisficing* que l'on peut comprendre au sens de faible implication du répondant est un phénomène que l'on peut appréhender à travers les modèles théoriques de réponse aux enquêtes. Dans ces modèles, les différentes étapes du processus cognitif qu'engage le répondant face à une question posée lors d'une enquête sont les suivantes : compréhension de la question, récupération de l'information pertinente, utilisation de cette information pour élaborer son jugement, sélection et report de la réponse (Tourangeau, 2000). Mais d'après la théorie du *satisficing* de Krosnick (1991), les répondants n'ont pas toujours la motivation ou l'attention nécessaires au bon déroulement du processus cognitif en 4 étapes. Ainsi, les réponses fournies ne sont pas toujours optimales. On parle plutôt de formes faibles de *satisficing* quand l'une au moins des étapes du processus cognitif n'est pas complètement optimale. Quand une étape du processus cognitif est sautée, on parle alors de formes fortes de *satisficing*.

Krosnick présente dans son article 6 formes différentes de *satisficing*. Une première forme, faible, consiste à choisir la première réponse qui semble satisfaisante et non celle qui est optimale. Ainsi, le choix de la réponse va fortement dépendre de la présentation orale ou visuelle des réponses, ce qui peut potentiellement donner lieu à des effets d'ordre de réponse (*primacy* pour la tendance à choisir ce qui est présenté en premier, *recency* pour ce qui est en dernier). Une autre forme faible de *satisficing* est selon Krosnick le fait d'acquiescer aux affirmations, pour les questions attitudinales, ou les questions d'opinion sur des politiques particulières.



Les 4 formes fortes de *satisficing* que Krosnick présente sont la tendance au choix du statu quo plutôt que le changement, la non-différenciation sur les échelles de réponse, le fait de répondre régulièrement "Ne sait pas" et enfin, le fait d'effectuer un choix aléatoire des réponses, une sorte de pile ou face mental pour sélectionner la réponse que l'on va reporter.

Selon Krosnick, la tendance au *satisficing* va dépendre de plusieurs facteurs. Un premier facteur est la difficulté de la question, que l'on peut aborder par le biais de la difficulté des 4 étapes du processus cognitif : certaines questions sont difficiles à interpréter, d'autres nécessitent un effort certain pour retrouver l'information pertinente, etc. Un autre facteur est celui de la capacité de l'enquêté à répondre sous deux aspects différents : d'un côté la facilité avec les opérations mentales plus ou moins complexes, de l'autre, la familiarité avec le sujet de l'enquête. Un dernier facteur est celui de la motivation intrinsèque ou extrinsèque à répondre. Ainsi, Krosnick exprime la tendance au *satisficing* par la formule (avec  $\alpha_1$ ,  $\alpha_2$  et  $\alpha_3$  fonctions croissantes) :

$$Satisficing = \frac{\alpha_1(Difficulté)}{\alpha_2(Capacité) \alpha_3(Motivation)}$$

### La notion de vraie valeur

Pour mesurer l'ampleur des erreurs systématiques, une valeur de référence est nécessaire, l'idéal absolu étant de disposer de la "vraie" valeur du paramètre d'intérêt. Mais, le concept de vraie valeur n'est pas toujours évident à appréhender. Ainsi, lorsqu'on s'intéresse à l'erreur de mesure, il y a l'idée d'un attribut permanent séparé de la mesure elle-même.

Les critères définissant la vraie valeur seraient les suivants (Hansen *et al.*, 1951) :

- elle doit être définie de façon unique ;
- elle doit être définie en fonction de l'objectif de l'enquête ;
- elle doit pouvoir être obtenue par des opérations concrètes, possibles à mener.

Pour Deming (1990), les vraies valeurs n'existent pas mais la valeur obtenue à partir d'une réalisation préférentielle de technique d'enquête peut être utilisée comme proxy. Ainsi, pour estimer des biais, on compare des estimations "ordinaires" à celles obtenues au moyen de procédures privilégiées qui ne peuvent pas être utilisées à grande échelle, dans un processus de type industriel.

Mais l'explication détaillée des difficultés concrètes des enquêtes conduit à s'éloigner, de fait, de cette idée de vraie valeur, en en gardant cependant le langage (Desrosières, 2012). D'un côté, le statisticien d'enquête essaie de répondre à une demande sociale de réalisme métrologique, en se calant sur le modèle des sciences de la nature, de l'autre, il ne peut négliger les médiations, les conventions qui régissent les phénomènes qu'il essaie de quantifier.

### 1.3.6 L'erreur de saisie, de codification

La saisie des données d'enquête est le processus inscrivant l'information communiquée par le répondant sur un support électronique. Cette étape est nécessaire dès lors que l'information n'est pas recueillie directement sous format électronique : c'est par exemple le cas pour une enquête où le mode de réponse est un questionnaire papier. L'étape de saisie peut être automatisée mais nécessite dans de nombreux cas une intervention humaine. Des erreurs peuvent

ainsi être commises lors de cette étape du processus d'enquête même lorsque des systèmes de double saisie ou des combinaisons de saisie optique et manuelle sont mis en oeuvre pour se prémunir de problèmes trop importants. Certes, l'importance de ce type d'erreur est sans doute à relativiser au regard de l'ensemble des composantes d'erreur d'enquête, mais cela peut néanmoins être problématique dès lors que les réponses sont complexes à saisir et les contrôles limités.

Certaines réponses d'enquête ne sont pas intégrées en tant que telles dans les tables de données mais subissent d'abord une codification. Cette étape de transformation peut être source d'erreur. En particulier des variables peuvent être codées via les réponses fournies à des questions ouvertes. Par exemple, à l'Insee, la codification des réponses sur la profession est automatisée grâce à un logiciel (Sicore, Système Informatisé de COdage des Réponses aux Enquêtes), mis à jour régulièrement. Ce programme de codification automatique utilise une base d'apprentissage pour transformer la réponse en code. Mais dans les cas où le logiciel ne parvient pas à classer la réponse, les reprises sont effectuées manuellement par des opérateurs. Cette reprise manuelle est une opération assez lourde, complexe et potentiellement source d'erreur.

## 1.4 Éléments de contexte historique

Très tôt, de nombreux auteurs ont remarqué qu'il existait bien d'autres erreurs d'enquête que celles attribuées à l'échantillonnage. Mais le concept d'erreur d'échantillonnage reste quand même central dans la littérature sur les enquêtes. Ainsi, les principaux textes sur la théorie des sondages insistent avant tout sur l'erreur d'échantillonnage : l'enquête par sondage pour assurer sa légitimité a dû se focaliser sur l'analyse de cette composante. Ceci est assez compréhensible car il était très important de pouvoir montrer qu'il était possible de réaliser des études statistiques en s'appuyant sur l'échantillonnage.

Et si de nombreuses typologies des erreurs existent, la distinction canonique est quand même de mettre d'un côté les erreurs liées à l'échantillonnage, de l'autre les erreurs non liées à l'échantillonnage. Cette distinction peut s'entendre d'un point de vue théorique comme pratique : si l'erreur d'échantillonnage peut être mesurée ou du moins estimée sur la plupart des échantillons probabilistes, les autres composantes de l'erreur d'enquête totale ne peuvent être mesurées directement, mais plutôt à partir de sources externes ou de protocoles alternatifs.

De nombreuses références ont défini et enrichi le concept d'erreur d'enquête totale, parmi lesquelles on peut citer par exemple :

- Deming (1944) liste les principales sources d'erreurs dans une enquête. Dans une liste de 13 sources d'erreurs, les erreurs d'échantillonnage sont citées mais n'occupent pas une place particulièrement prépondérante. En revanche, les erreurs de couverture ne sont pas mentionnées.
- Kish (1965) consacre 65 des 643 pages de son livre *Survey Sampling* aux erreurs autres que l'échantillonnage. En particulier, il insiste sur les biais et va distinguer les erreurs d'observation des autres types d'erreur.
- Dalenius (1974) se place dans un cadre plus global du *design* d'enquête et tient compte

à la fois des aspects méthodologiques, financiers et informationnels.

- Anderson, Kasper et Frankel (1979) vont consacrer l'expression d'erreur d'enquête totale puisque leur livre s'intitule *Total Survey Error* : néanmoins il ne s'agit pas d'un livre présentant uniquement un corpus théorique mais également un recueil d'études empiriques à partir de données d'enquêtes sur la santé traitant entre autres de problèmes de non-réponse, de mesure, etc. Trois typologies des composantes d'erreur sont présentées : biais/variance, échantillonnage/non-échantillonnage et observation/non-observation.
- Groves (1989) enrichit le concept avec des éléments provenant de la psychométrie. Par exemple il discute des liens entre les notions de validité de la psychométrie et les concepts de méthodologie d'enquête.
- Lessler et Kalsbeek (1992) effectuent dans leur livre *Nonsampling Error in Surveys* une synthèse sur les trois principales erreurs d'enquête autres que l'échantillonnage : couverture, non-réponse et mesure ; plus récemment McNabb (2014) traite également des erreurs d'enquête autres que l'échantillonnage dans le contexte des enquêtes sociales.
- Biemer et Lyberg (2003) intègrent des éléments de qualité des processus, notions utilisées dans le domaine de l'industrie.
- Weisberg (2009) fournit des éléments théoriques sur les différentes composantes de l'erreur d'enquête ainsi que les solutions potentielles pour limiter l'erreur.

## 1.5 Les apports du concept

Dans leur article *Can a Statistician Deliver*, Särndal et Platek (2001) font une synthèse sur les progrès et les échecs dans le domaine de la méthodologie d'enquête. Ils insistent en particulier sur les différentes catégories de statisticiens impliqués, qui de par leurs différences peuvent avoir des opinions divergentes sur ce que doit être la qualité statistique. Särndal et Platek regrettent ainsi que "Quand le statisticien d'enquête parle de précision, il se réfère surtout à la seule erreur d'échantillonnage." Un apport important du concept d'erreur d'enquête totale est de bien distinguer les différentes composantes de l'erreur d'enquête et de ne pas se restreindre à la seule erreur d'échantillonnage.

Le cadre de l'erreur d'enquête totale fournit une décomposition des erreurs, sépare la variance du biais, l'observation de la non-observation, tout ceci en détaillant bien les différentes étapes du processus d'enquête.

Il s'agit d'un cadre particulièrement bien adapté pour comparer différents protocoles de collecte ou encore envisager la bascule d'un mode de collecte à un autre.

Certains domaines restent encore un peu sous-étudiés (comme les erreurs de spécification ou les erreurs de saisie) mais dans l'ensemble, l'élargissement des connaissances sur les erreurs d'enquête suscite un véritable intérêt.

## 1.6 Limites et extensions

### 1.6.1 Précision ou qualité ?

Le cadre de l'erreur d'enquête totale permet d'analyser la précision au sens large, mais qu'en est-il de la qualité de l'enquête ? La précision d'une enquête n'est pas toujours suffisante pour évaluer la qualité d'une enquête qui s'insère dans un contexte bien particulier. En particulier, le concept d'erreur d'enquête totale ne tient pas vraiment compte de la dimension utilisateur, à savoir dans quelle mesure l'enquête répond aux attentes et aux besoins des utilisateurs ? D'après Lyberg (2012), la qualité des enquêtes est un concept multidimensionnel issu de deux démarches de développement distinctes. La première démarche s'inscrit dans le paradigme de l'erreur d'enquête totale. La deuxième démarche repose plutôt sur des concepts des sciences de la gestion de la qualité.

### 1.6.2 La qualité totale d'enquête

L'erreur d'enquête totale peut s'insérer dans un cadre plus général dit de la qualité d'enquête totale, en anglais *Total Survey Quality* (TSQ) où la TSE ne serait qu'une composante, en tant qu'indicateur de précision. (Biemer et Lyberg, 2003)

Les cadres de qualité de certains organismes de statistique publique retiennent des critères bien particuliers que l'on peut rapprocher des différentes dimensions de la TSQ. Ainsi, on peut considérer que les principales dimensions de la TSQ sont les suivantes :

- précision ;
- pertinence ;
- actualité ;
- accessibilité ;
- cohérence ;
- comparabilité.

Ces dimensions sont les six critères de qualité énoncés par Eurostat (Desrosières, 2003). Ces critères sont largement repris dans le code de bonnes pratiques de la statistique européenne de 2005<sup>2</sup> où ils sont énoncés en tant que principes devant s'appliquer aux résultats mis à disposition.

La précision porte sur l'écart entre l'estimation à partir des données d'enquête et la valeur réelle, inconnue, du paramètre d'intérêt. Un indicateur pertinent de cette précision est l'erreur d'enquête totale.

La pertinence renvoie à la satisfaction des besoins des utilisateurs. L'identification de ceux-ci et de leurs attentes est dès lors un point essentiel.

Par actualité (ou rapidité, fraîcheur), on entend le délai entre le phénomène décrit par l'enquête et la disponibilité des données. Ce délai ne doit pas être trop long faute de quoi les résultats de l'enquête risquent de ne pas pouvoir éclairer le débat public.

L'accessibilité d'une enquête est la facilité avec laquelle les utilisateurs peuvent récupérer les données d'enquête puis les exploiter. Outre la localisation, le format des données, leur documentation sont quelques uns des facteurs d'accessibilité (ou de clarté des données d'enquête).

---

2. Plus précisément, dans la partie "Résultats disponibles" du code de bonnes pratiques, 5 principes sont énoncés : pertinence, exactitude et fiabilité, actualité et ponctualité, cohérence et comparabilité, accessibilité et clarté.

La cohérence reflète la mesure dans laquelle les produits de données sont reliés logiquement et mutuellement concordants.

Un dernier critère est celui de la comparabilité des données, à la fois dans le temps et dans l'espace.

D'autres dimensions "utilisateurs" peuvent être intégrées comme par exemple la crédibilité qui est un critère retenu par l'OCDE dans son cadre de qualité.

Optimiser la qualité d'enquête totale nécessite que les objectifs de qualité pour chaque dimension soient bien définis et que les approches pour atteindre ces objectifs soient optimales et budgétées. Dans l'approche de Biemer et Lyberg (2003), l'objectif est de maximiser la précision des données sous des contraintes de dimensions "utilisateurs".

Si la précision est insuffisante, les autres dimensions sont sans pertinence, mais l'inverse est vrai également : des données très précises peuvent être inutiles si diffusées trop tard pour avoir une incidence sur les décisions. Par ailleurs, les dimensions de la qualité sont souvent conflictuelles : actualité vs précision ou encore comparabilité vs précision.

## 1.7 Panorama des questions de recherche

L'intérêt des méthodologues est grandissant concernant le concept d'erreur d'enquête totale. En 2005 est organisé un séminaire (*International Total Survey Workshop*) où les chercheurs présentent les dernières avancées sur l'erreur d'enquête totale, qui prend un rythme annuel à partir de 2008. Le corpus théorique des erreurs non liées à l'échantillonnage est, par nature, moins cohérent que celui de la théorie des sondages et l'accumulation des expériences est nécessaire pour mieux connaître et évaluer les différentes composantes d'erreur et leurs éventuelles interactions, ceci afin d'accroître la précision, au sens large, des enquêtes. Parmi les nombreux thèmes abordés dans la littérature, nous choisissons d'en présenter quelques uns.

### 1.7.1 Les effets enquêteurs

La question est de savoir dans quelle mesure l'enquêteur a un impact sur les réponses fournies par l'individu enquêté ou leur qualité. Cet impact peut être lié à la manière dont il administre l'enquête, lit les questions, fournit des instructions (*role-dependent effect*) (Brenner, 1982), mais aussi aux caractéristiques sociodémographiques de l'enquêteur perçues par le répondant (*role-independent effect*) (Hagenaars et Heinen, 1982). Par exemple, Davis *et al.* (2010) font une revue de littérature sur ces questions d'effet enquêteur lié aux caractéristiques sociodémographiques sur les enquêtes en santé publique.

Comme exemple d'autres travaux sur les enquêteurs et l'erreur d'enquête totale, nous pouvons citer ceux de Yan *et al.* (2008) qui s'intéressent à la tâche duale des enquêteurs : d'un côté, recruter des répondants potentiels, de l'autre administrer les enquêtes, c'est-à-dire, lire les questions, recueillir les réponses, fournir des instructions, etc. L'étape de recrutement a des conséquences plutôt sur l'erreur de non-réponse, celle d'administration sur l'erreur de mesure. La question que se posent les auteurs est celle de l'existence ou non d'une relation dans la

performance des enquêteurs sur ces deux étapes de leur travail.

### 1.7.2 Protocoles d'enquêtes adaptatifs

Une enquête a beau être préparée le plus soigneusement possible, elle n'est pas à l'abri de dysfonctionnements plus ou moins majeurs sur le terrain, ou de chocs difficilement prévisibles. Les protocoles d'enquêtes adaptatifs sont des protocoles qui ont la capacité de s'adapter aux difficultés rencontrées sur le terrain sur la base des premières remontées sur des indicateurs clés de la qualité. Certains modes de collecte sont plus indiqués que d'autres pour ce type de protocoles. Ainsi, la collecte assistée par ordinateur (Capi ou Cawi par exemple) permet une surveillance quasiment en temps réel de la qualité des données. Des données partiellement collectées ou des parodonnées peuvent être ainsi utilisées pour modifier la stratégie de collecte. Par exemple, on peut envisager de donner de nouvelles formations aux enquêteurs sur la passation du questionnaire, si l'on constate que la qualité des réponses fournies n'est pas celle attendue (Biemer et Lyberg, 2003).

Un autre exemple est celui d'essayer de réduire le biais de non-réponse en ciblant mieux les efforts de l'organisation enquêtrice. Pour cela, on utilise les données et les parodonnées remontant d'une première phase de la collecte (Peytchev *et al.*, 2010). Pour des développements théoriques sur ce sujet, on pourra se référer à (Schouten *et al.*, 2011) ou à (Beaumont et Haziza, 2011). Par ailleurs, toute une littérature existe sur l'utilisation des parodonnées dans les processus d'enquêtes, et en particulier dans le cadre d'analyse de l'erreur d'enquête totale (Kreuter, 2013).

### 1.7.3 Comparaison des différents modes de collecte

Nous verrons plus en détail, dans le chapitre sur la collecte multimode, des exemples de comparaison entre les différents modes de collecte. L'idée est de voir s'il y a un effet mode de collecte dans les réponses recueillies. Les principaux modes qui font l'objet de ces comparaisons sont le face à face, le téléphone, le papier et internet. Ces comparaisons portent donc plutôt sur les erreurs de mesure et de nombreuses questions méthodologiques se posent dès lors que l'on souhaite obtenir un effet mode "pur" relatif à la mesure, purgé des autres biais éventuels.

### 1.7.4 Arbitrages erreur de couverture, erreur de non-réponse

Pour Groves et Lyberg, un des principaux axes de la recherche future sur l'erreur d'enquête totale est celui des interactions entre les différentes composantes d'erreur (Groves et Lyberg, 2010).

Par exemple, Eckman (2012) traite de l'arbitrage entre les erreurs de couverture et les erreurs de non-réponse. Ainsi, les efforts pour améliorer la couverture peuvent être divers, suivant la nature et les objectifs de l'enquête : faire en sorte que les personnes ayant uniquement un téléphone mobile puissent être atteints quand la base de sondage initiale repose sur des numéros de téléphone fixe, fournir un accès à internet et des ordinateurs aux personnes qui n'en ont pas, etc. Le problème est, selon l'auteur, que ces efforts coûteux peuvent être en partie vains quand les personnes nouvellement couvertes sont beaucoup plus non répondantes que celles de la population initiale. L'auteur se demande ainsi si un protocole avec un faible taux de réponse

et une bonne couverture est préférable ou non à un protocole avec un meilleur taux de réponse mais une plus faible couverture. Pour répondre à cette question, il faut en revenir aux biais de couverture et de non-réponse et leurs interactions : comment ces deux types de biais, les propensions à répondre et à être couvert sont reliés entre eux ?

### 1.7.5 Interactions entre erreurs de non-réponse et erreurs de mesure

D'autres exemples d'interactions entre composantes d'erreur sont étudiés dans la littérature, en particulier ceux entre les erreurs de non-réponse et les erreurs de mesure. En effet, si les efforts pour réduire le biais de non-réponse se traduisent également par des erreurs de mesure plus importantes, cela peut être contre-productif du point de vue de l'erreur d'enquête totale (Groves et Lyberg, 2010). Olson fait une synthèse des hypothèses qui relient la décision de participation à l'erreur de mesure : ces hypothèses sont liées soit à la motivation, l'intérêt pour le thème, la sensibilité du thème ou encore les caractéristiques individuelles de l'individu (Olson, 2008). Par exemple la motivation est un facteur qui affecte la propension à répondre tout comme la qualité de la réponse. Ceux qui ont été de prime abord plutôt réticents à participer ont une certaine tendance au *satisficing*. Kreuter *et al.* (2010) étudient cette question en utilisant des données et par données d'une enquête sur la participation au marché du travail, appariées à des données administratives. Ils trouvent qu'augmenter les tentatives de contact réduit de façon significative le biais de non-réponse, mais cet effort augmente l'erreur de mesure au point que le biais total est à peine réduit.

## Chapitre 2

# La collecte multimode à travers le prisme de l'erreur d'enquête totale

La plupart des instituts de statistique s'engagent ou se sont engagés récemment dans des processus de refonte de leurs enquêtes ménages. Un des axes de ces programmes de refonte est d'utiliser de nouveaux moyens de communications pour atteindre l'individu que l'on souhaite enquêter. Ainsi, les enquêtes et même plus précisément les collectes prennent de plus en plus des formes multimodes, par exemple pour répondre à des questions de couverture de la population cible. En particulier, internet est probablement le mode qui est amené à se développer ces prochaines années mais son introduction dans les processus de production doit se faire avec précaution compte tenu des nombreuses questions qui existent.

Un cadre particulièrement adapté pour évaluer les modifications ou les évolutions des protocoles d'enquête est celui de l'erreur d'enquête totale. En effet, réfléchir au sein de ce paradigme permet d'évaluer les conséquences des modifications de protocole d'un point de vue global. Par exemple, si l'on introduit un nouveau mode de collecte essentiellement pour des questions de couverture, il faut également avoir une idée des conséquences potentielles sur l'erreur de non réponse ou sur l'erreur de mesure. Certes, il ne sera sans doute pas toujours facile de quantifier de manière exacte les impacts de modifications de protocoles sur les différentes composantes de l'erreur d'enquête, mais il est important d'avoir cette démarche pour mesurer la conséquence globale sur l'erreur d'enquête, au moins en première approximation, pour justifier de la pertinence du nouveau protocole mis en oeuvre.

### 2.1 Un peu de vocabulaire

Tout d'abord, essayons de définir les termes qui vont souvent revenir dans ce chapitre.

Par mode, on entend façon de faire (Amiel et Denoyelle, 2012). Il se définit aux différents stades du processus de l'enquête. Ainsi, le mode de contact est le moyen utilisé pour contacter les enquêtés, par exemple une lettre-avis envoyée par courrier. Le mode de collecte est la façon de recueillir l'information auprès de l'enquêté. Le mode de suivi est le moyen utilisé pour les contacts ultérieurs à la phase de collecte principale.

Une enquête est dite multimode (en anglais *multi-mode survey* ou *mixed-mode survey*)



lorsque différents canaux de communication sont intégrés au processus d'enquête. C'est le cas par exemple lorsqu'un courrier est envoyé au préalable au ménage pour annoncer la visite d'un enquêteur qui interrogera en face à face. Si l'utilisation du terme multimode est somme toute assez récente, de très nombreuses enquêtes le sont déjà.

La collecte est multimode quand différents modes de communication sont intégrés plus précisément dans la phase de collecte. Par exemple, quand une partie de l'échantillon sélectionné est interrogé en face à face, l'autre partie par téléphone. L'enquête est également multimode quand des individus sont d'abord interrogés selon un mode de collecte puis réinterrogés plus tard à l'aide d'un autre mode de collecte. C'est par exemple le cas de l'enquête Emploi de l'Insee où les premières et dernières interrogations se font en face à face alors que les interrogations intermédiaires se font par téléphone.

D'après Dillman, on peut distinguer trois principales formes de collecte multimode (Dillman, 2009) :

- le multimode intégré : chaque répondant d'une enquête est interrogé via plusieurs modes, par exemple une partie d'un questionnaire administrée par un enquêteur, l'autre partie auto-administrée, souvent pour réduire les biais de désirabilité sociale ou pour des questions de confidentialité ;
- le multimode concurrent : plusieurs modes de réponse sont offerts ou affectés aux enquêtés, principalement pour réduire les erreurs de couverture et de non-réponse ;
- le multimode séquentiel : différents modes de réponse proposés aux différentes étapes du processus de collecte, par exemple un mode proposé lors d'une première phase, un mode alternatif pour les non-répondants de la première phase, là aussi pour améliorer réponse et couverture.

Les principaux modes de collecte sont le face à face, le téléphone, le papier et internet. Lorsque l'interrogation en face à face est assistée d'un ordinateur, on parle de Capi (*computer assisted personal interviewing*) et pour le téléphone et internet, respectivement de Cati (*computer assisted telephone interviewing*) et Cawi (*computer assisted web interviewing*). Aujourd'hui, à l'Insee, pour les enquêtes ménages, l'essentiel des interrogations se font par Capi ou par Cati, Capi étant largement majoritaire si les parts respectives de chaque mode sont calculées sur la base de la durée de passation des questionnaires<sup>1</sup>. D'autres modes de collecte, un peu moins utilisés, existent. Par exemple, lorsque l'individu reporte lui-même ses réponses sur un ordinateur, on parle de CASI (*computer assisted self-interviewing*). L'enquête Cadre de Vie et Sécurité de l'Insee utilise pour les questions les plus sensibles un protocole sous casque, on parle alors d'audio-CASI. Enfin, on peut également citer l'utilisation de serveurs vocaux interactifs dans certains instituts de statistique ; les réponses sont données via le clavier téléphonique (TDE pour *touchtone data entry*) ou oralement interprétées grâce à la reconnaissance automatique de la parole (ASR pour *automatic speech recognition*).

---

1. La répartition entre les deux modes est un peu plus équilibrée sur la base du nombre de questionnaires réalisés ; le téléphone est essentiellement utilisé pour les nombreuses, mais courtes, ré-interrogations intermédiaires de l'enquête Emploi.

## 2.2 Pourquoi le multimode ?

Intégrer plusieurs modes de collecte au sein des protocoles d'enquête est une des solutions que l'on peut apporter pour enrayer la baisse tendancielle des taux de réponse. En effet, il s'est produit au fil du temps une sorte de renversement de la balance du pouvoir, qui est dorénavant du côté des enquêtés (Dillman, 2009). Ces derniers contrôlent les contacts et sélectionnent les entités à qui ils veulent bien répondre<sup>2</sup>. Pour les instituts de statistique, la question est d'adapter les protocoles d'enquête pour essayer d'atteindre les personnes qui ne répondent plus aux enquêtes. La collecte multimode peut être alors le moyen de mieux s'adapter aux préférences des individus sur la manière de répondre à une enquête. Par exemple, si les taux de réponse des enquêtes utilisant internet comme mode de collecte sont généralement moins élevés que ceux des enquêtes en face à face, il y a sans doute des personnes qui répondent sur internet mais ne le font pas en face à face. Millar *et al.* (2009) ont par exemple étudié les préférences des individus sur les modes de collecte : celles-ci sont fortement corrélées aux caractéristiques des individus, en particulier, l'âge, le niveau d'éducation et les revenus.

L'un des enjeux importants est celui des coûts. Passer à la collecte multimode en introduisant des modes de collecte peu coûteux, comme internet, mode dont le coût marginal d'un questionnaire est quasiment nul, permettrait de faire des économies significatives. Les méthodologues sont bien conscients de ces enjeux mais reste qu'il est important, malgré les pressions qui peuvent exister, d'étudier précisément les conséquences de ces évolutions du point de vue de la qualité des données avant toute velléité de généralisation.

Le recours au multimode peut d'ailleurs s'appréhender, de manière un peu caricaturale, selon deux approches (Schouten et van Dam, 2013) : une approche par les coûts, où le multimode permet de diminuer les coûts en maintenant la qualité (plutôt l'approche de l'institut de statistique) et une approche par la qualité où l'objectif est d'améliorer la qualité sans augmenter les coûts (plutôt l'approche de la recherche académique).

Enfin, si l'on se place dans le cadre du paradigme de l'erreur d'enquête totale, une collecte multimode pourrait être une solution optimale. En effet, l'idée est que le multimode permet de compenser les faiblesses des différents modes de collecte. Certains modes sont déficients du point de vue de la couverture, d'autres ont leur faiblesse côté erreur de mesure ou sur le biais de non-réponse, si bien qu'une solution multimode est dans certains cas le meilleur moyen de minimiser l'erreur d'enquête totale.

## 2.3 Multimode et TSE

### 2.3.1 TSE : un cadre adapté ?

Dès lors que l'on s'interroge sur la pertinence d'un protocole un peu complexe, il semble naturel de se placer dans un cadre qui va au-delà de l'erreur d'échantillonnage, qui tient compte

---

2. Pour certaines organisations enquêtrices, le "renversement" est tel qu'elles envisagent dorénavant de rémunérer les enquêtés.

de toutes les erreurs aux différentes étapes du processus d'enquête. Ainsi, la tendance à aller vers des collectes multimodes relance en quelque sorte l'intérêt sur l'erreur d'enquête totale et sa décomposition en erreurs ; ceci d'autant plus que les modes de collecte se différencient pour certains significativement sur les différentes composantes. Le cadre de l'erreur d'enquête totale s'avère particulièrement adapté pour comparer différentes possibilités de protocoles, préparer des changements dans les protocoles ou encore tester des protocoles innovants.

### 2.3.2 Composantes d'erreur et solutions multimodes

D'un point de vue théorique, le protocole d'enquête optimal est celui qui maximise la précision et donc minimise l'erreur d'enquête totale. Si l'on passe les différentes composantes une par une, il existe des solutions multimodes pour diminuer chacune des erreurs. Il faut quand même garder à l'esprit que modifier un protocole pour diminuer une composante d'erreur peut avoir des conséquences sur les autres composantes avec le risque que cela augmente au final l'erreur totale.

Pour réduire l'erreur de couverture d'une enquête par internet, on peut combiner ce mode à faible coût marginal mais ne couvrant pas toute la population avec un autre mode moins sujet aux problèmes de couverture (multimode concurrent).

Pour réduire l'erreur de non-réponse d'une enquête, on peut, dans un deuxième temps, offrir un mode alternatif de réponse aux non-répondants (multimode séquentiel).

Pour réduire l'erreur de mesure d'une enquête abordant des questions sensibles, on peut utiliser un mode pour l'essentiel du questionnaire et un mode différent (auto-administré) pour le sous-ensemble des questions sensibles (multimode intégré).

## 2.4 Comparaison entre les modes

Les modes de collecte se différencient par de nombreux aspects. En particulier, les principales composantes de l'erreur d'enquête totale non liées à l'échantillonnage que sont les erreurs de couverture, de non-réponse et de mesure vont dépendre du mode de collecte. Par ailleurs, les coûts associés aux différents mode de collecte ne sont pas les mêmes. C'est d'ailleurs l'une des raisons mêmes qui pousse au développement de la collecte multimode : l'objectif étant souvent d'intégrer des modes de collecte moins coûteux tout en préservant la qualité des données recueillies.

### 2.4.1 Couverture

On peut distinguer de manière plus ou moins artificielle les problèmes de couverture sur deux niveaux différents. Le premier niveau est celui de la base de sondage ; dans quelle mesure la base de sondage utilisée couvre théoriquement toute la population ? L'autre niveau est celui de la capacité à répondre pour les personnes sélectionnées : est-ce que ces personnes peuvent répondre avec le protocole qui leur est proposé ?

Ici on va plutôt s'intéresser aux modes de collecte, indépendamment de la base de sondage. En

effet, il n'y a pas forcément correspondance entre la base de sondage, le mode de contact et le mode de collecte. Pour une enquête purement par voie électronique on pourrait ainsi utiliser une base de sondage constituée des adresses électroniques des personnes de la population cible, les contacter par courriel et leur demander de répondre par internet. Ce genre d'enquête peut être réalisé au sein d'entreprises où tout le monde disposerait d'une adresse électronique. Mais ce n'est aujourd'hui pas possible de mettre en place ce type de protocole pour enquêter en population générale. Par exemple, à l'Insee, lorsque les ménages sont enquêtés par internet, le contact est établi en envoyant un courrier à une adresse postale provenant de la base de sondage. Pour la réponse en face à face, par téléphone ou par papier, dès lors que le contact est établi, la réponse est dans une très large majorité des cas possible. Néanmoins, il ne faut pas occulter que dans certains cas, la personne ne pourra pas répondre (illettrisme, handicaps physiques). En ce qui concerne internet, le problème de couverture est bien moins négligeable : tout le monde n'a pas encore accès à ce moyen de communication, même si cet accès s'est largement démocratisé ces dernières années puisqu'en 2012, 75 % des personnes ont utilisé internet récemment (au moins une fois au cours des trois derniers mois) alors que ce taux était de 57 % en 2007 (Gombault, 2013). Une partie de la population française ne peut donc néanmoins toujours pas répondre sur internet. Et cette population possède des caractéristiques sociodémographiques différentes et donc des valeurs sur les variables d'intérêt potentiellement différentes, d'où l'existence d'un biais de couverture, qui peut malgré tout être corrigé, au moins partiellement.

en %

	Individus disposant d'un micro-ordinateur à la maison	Individus disposant d'Internet à la maison	Fréquence de l'utilisation d'Internet		Individus ayant utilisé l'Internet mobile au cours des 3 derniers mois
			Au cours des 3 derniers mois (Internaute)	dont tous les jours ou presque	
<b>Ensemble des répondants</b>	<b>79,1</b>	<b>78,2</b>	<b>75,0</b>	<b>79,7</b>	<b>39,7</b>
Homme	82,4	81,5	77,9	81,6	45,0
Femme	76,0	75,2	72,3	77,8	34,8
15-29 ans	96,4	96,2	97,7	88,3	75,0
30-44 ans	92,6	91,9	92,2	79,4	50,8
45-59 ans	87,8	86,5	82,3	75,9	33,7
60-74 ans	63,0	61,9	52,2	74,3	16,4
75 ans et plus	24,4	23,4	16,5	60,9	3,1
<b>Actifs occupés</b>	<b>92,2</b>	<b>91,6</b>	<b>90,6</b>	<b>80,4</b>	<b>50,1</b>
Agriculteurs	75,1	70,6	62,6	72,2	17,8
Artisans et commerçants	92,1	90,5	85,6	77,7	52,6
Cadres et prof. libérales	98,4	97,6	99,1	92,3	70,9
Professions intermédiaires	95,8	95,0	97,1	86,6	56,3
Employés	89,7	89,6	88,0	72,9	39,1
Ouvriers	88,2	88,2	84,1	72,0	41,5
<b>Chômeurs</b>	<b>86,3</b>	<b>85,7</b>	<b>87,0</b>	<b>81,0</b>	<b>47,1</b>
<b>Etudiants</b>	<b>98,3</b>	<b>97,9</b>	<b>99,1</b>	<b>89,7</b>	<b>78,1</b>
<b>Retraités</b>	<b>50,3</b>	<b>49,2</b>	<b>40,9</b>	<b>72,0</b>	<b>11,7</b>
Sans diplôme ou certificat d'études primaires	54,1	53,2	43,6	65,4	16,6
BEP, CAP, BEPC	82,8	82,6	79,2	75,6	38,5
Baccalauréat	91,4	90,5	91,3	83,0	50,6
Diplôme d'études supérieures	96,0	94,5	96,5	90,1	62,3

Lecture : 97,7 % des 15-29 ans ont utilisé Internet au cours des 3 derniers mois, dont 88,3 % tous les jours ou presque.  
 Champ : personnes de 15 ans ou plus vivant en France métropolitaine dans un ménage ordinaire.  
 Source : Insee, enquête Technologies de l'information et de la communication 2012.

FIGURE 2.1 – Accès et usage d'internet en 2012 (Gombault, 2013)

## 2.4.2 Taux de réponse

Tous les modes de collecte ne sont pas aussi efficaces les uns que les autres pour obtenir la participation des enquêtés. Les comparaisons des taux de réponse entre les différents mode de collecte reposent le plus souvent sur des méta-analyses (Groves et Peytcheva, 2008). De ces méta-analyses, il ressort que le face à face est le mode de collecte qui génère les taux de réponse les plus élevés suivi par le téléphone, le courrier et enfin le web. Mais au-delà de cette hiérarchie générale, la comparaison n'est pas si aisée car les écarts entre taux de réponse peuvent refléter différents facteurs : mode de contact, relances, personnalisation ou encore incitations à répondre.

Entre le face à face et le téléphone, les différences s'expliquent principalement par le rôle de l'enquêteur : sa capacité à persuader est moins forte quand il ne peut jouer que par le canal auditif. Et faute d'enquêteur pour persuader et assurer de la légitimité de l'enquête, il n'est pas surprenant que les questionnaires auto-administrés génèrent moins de réponses : la participation est intrinsèquement liée à l'intérêt pour le sujet de l'enquête. Une raison avancée pour expliquée la supériorité du papier sur le web par Millar et Dillman (2011) est que la vue générale du questionnaire papier est plus susceptible de pousser à la participation que la version web. Pour le questionnaire web, il faut faire l'effort d'aller s'installer sur son ordinateur sans trop savoir vers quoi on s'engage en matière de temps, de thèmes ou de difficulté des questions, informations qui ne sont dévoilées qu'au fur et à mesure du questionnaire, puisque les questions sont souvent posées une par une. Sur papier, un rapide coup d'oeil permet de rapidement se décider à participer ou non. En revanche, la non-réponse partielle est plus fréquente sur papier que sur internet où les possibilités d'implémenter des contrôles sont nombreuses (contrôles bloquants pour les variables clés du questionnaires, affichage de la modalité de non réponse uniquement dans un deuxième temps, etc.).

Un exemple d'enquête utilisant une collecte multimode dans son processus de production aujourd'hui à l'Insee est l'enquête TIC déjà évoquée dans ce document. Pour les individus sélectionnés, retrouvés dans l'annuaire et donc interrogés par téléphone, le taux de réponse est de 74 %. Les individus qui ne sont pas retrouvés dans l'annuaire peuvent répondre soit par internet soit par papier : le taux de réponse global est de 36 % avec des réponses qui se répartissent à peu près équitablement entre les deux modes (Gombault et Duée, 2012). Un autre exemple d'introduction d'internet dans le processus de collecte pour améliorer le taux de réponse auprès d'une population particulière est celui du Baromètre santé médecins généralistes 2009 : une telle introduction pouvait être envisagée étant donné l'informatisation généralisée des médecins de ville (Gautier *et al.*, 2011).

## 2.4.3 Erreur de mesure

Sur l'erreur de mesure, les modes de collecte se distinguent essentiellement par la dichotomie entre les modes auto-administrés et les modes avec enquêteurs. D'un côté, le biais de désirabilité sociale est *a priori* plus fort quand l'enquêteur est présent. De l'autre, en l'absence d'enquêteur la tendance au *satisficing* est plus importante, car l'effort cognitif demandé est plus important. D'un côté, l'enquêteur peut aider à la bonne compréhension des questions, de l'autre, il peut influencer les réponses. Du point de vue de l'erreur de mesure, il est ainsi plutôt difficile d'établir une hiérarchie claire entre ces deux modes d'administration des enquêtes.

Entre face à face et téléphone, l'avantage est plutôt au face à face : le *satisficing* y est moindre, car l'enquêté peut être aidé par la communication non verbale et la pression temporelle (rythme de l'interview) est généralement moins forte. En revanche, les biais de désirabilité sociale seraient *a priori* moins forts au téléphone qu'en face à face car la distance sociale est plus importante. Néanmoins quelques études empiriques nuancent cela ; pour certains auteurs, les biais de désirabilité sociale serait équivalents (Beukenhorst et Wetzels, 2009) voire plus forts au téléphone qu'en face à face (Jäckle *et al.*, 2010; Holbrook *et al.*, 2003).

Les internautes sont des habitués du multi-tâches (Watson, 2009) : ils peuvent lancer plusieurs applications simultanément sur leur ordinateur, ils peuvent ouvrir plusieurs fenêtres ou onglets de navigation web. Ainsi les répondants sur internet ont plus tendance au *satisficing* que les répondants papier. D'ailleurs un autre élément qui va dans cette direction est que la façon de lire des internautes diffère quelque peu : beaucoup plus que la lecture sur papier, celle sur internet se fait en parcourant rapidement l'écran à la recherche de mots ou de phrases clés.

Notons que l'un des gros avantages de l'entretien réalisé avec déplacement d'un enquêteur et qu'il permet de combiner, dans une même séquence et avec une certaine fluidité, plusieurs approches : passation en face à face, questionnaire auto-administré sur papier ou sur ordinateur, éventuellement avec un casque pour les questions les plus sensibles (audio-CASI), etc.

#### 2.4.4 Coûts

Sur les coûts, la hiérarchie entre les 4 grands modes de collecte est bien établie : le face à face est le mode le plus coûteux, suivi par le téléphone, le papier et enfin internet. Des ordres de grandeur donnés dans la littérature sont que le face à face est environ 2 fois plus cher que le téléphone, lui-même deux fois plus cher que le papier (Roberts, 2007). Internet est quant à lui moins coûteux que le papier puisque les coûts d'impression et d'affranchissement sont réduits. Pour certaines enquêtes, ils ne sont pas forcément nuls puisque la notification préalable peut se faire par courrier, ainsi que les relances. Un des autres gros avantages en termes de coût d'internet par rapport au papier est que les traitements post-collecte nécessaires sont bien moins importants. En particulier, il n'y a pas besoin de saisir les questionnaires ce qui engendre une baisse de coût non négligeable. L'apurement des données d'une enquête par internet est également plus aisé que pour une enquête papier puisque les contrôles intégrés au questionnaire permettent de limiter le volume des données à corriger *a posteriori*.

#### 2.4.5 La collecte par internet

Internet s'est largement installé dans le quotidien des personnes et c'est tout naturellement que les instituts de statistique considèrent ce moyen de communication pour recueillir l'information auprès des populations cibles. Et si les bonnes pratiques de conception des questionnaires en ligne ne sont pas encore définitivement arrêtées, la littérature sur les questions d'ergonomie ou de design des questionnaires est déjà très riche (Couper, 2008).

Internet peut être utilisé comme mode unique pour certaines enquêtes mais sera le plus souvent un mode complémentaire à d'autres modes de collecte : la volonté de recourir à ce mode de collecte pour les avantages qu'il offre explique pour partie le regain d'intérêt de la recherche en méthodologie d'enquête sur la collecte multimode. Néanmoins, une rapide évaluation de ce mode de collecte à travers le prisme de l'erreur d'enquête totale montre qu'il est, aujourd'hui,

loin d'être exempt de défauts.

### Avantages de la collecte par internet

Un avantage indéniable de la collecte par internet est sa rapidité. De nature électronique, elle s'intègre naturellement dans des processus de production à grande échelle et sa flexibilité permet d'envisager des enquêtes en population générale dans des délais assez serrés. Lorsque les moyens informatiques sont bien calibrés, il n'y a pas de goulet d'étranglement durant la collecte comme cela peut exister pour les autres modes en raison du nombre limité d'enquêteurs. Et alors qu'une collecte de questionnaires papier auto-administrés implique des délais d'envois postaux et de saisie des questionnaires, l'accessibilité des données est potentiellement quasi immédiate lors d'une collecte par internet.

Si les instituts envisagent d'intégrer internet dans leurs processus de collecte, c'est aussi parce qu'il s'agit d'un mode peu coûteux, beaucoup moins cher que le face à face par exemple mais aussi moins cher que le téléphone et dans une moindre mesure que le papier. Le coût marginal d'une réponse par internet est d'ailleurs quasi-nul, ce qui permet d'envisager de très grandes tailles d'échantillons.

Le problème d'un questionnaire auto-administré sur papier est qu'il laisse le répondant un peu seul dans son exercice cognitif. Internet permet en partie d'alléger le fardeau du répondant puisque l'un des avantages de ce mode de collecte est son interactivité. Il s'agit d'un mode dynamique qui sollicite le répondant en cas d'incohérence des réponses, indique la progression dans le questionnaire ou encore fournit des instructions ou des précisions (par le biais d'info-bulles par exemple) pour l'aider à remplir le questionnaire.

Et là aussi, si on le compare avec la version papier, le questionnaire en ligne possède un net avantage en termes de complexité. Alors que les filtres doivent être utilisés avec parcimonie sur papier, internet offre beaucoup de possibilités à ce niveau-là. Le questionnaire peut théoriquement être très riche en termes de contenu : image, son ou encore vidéo sont *a priori* possibles.

Un autre avantage de la collecte par internet est que le choix du moment du remplissage est à discrétion du répondant. Ce mode de collecte a un caractère peu intrusif, peu contraignant si on le compare aux modes qui font intervenir un enquêteur. Par internet, l'enquêté est libre de répondre à un horaire qui lui convient, il s'agit ainsi d'un mode particulièrement bien adapté pour les personnes difficilement joignables en journée.

### Mais ...

Aujourd'hui, certains segments de la population n'ont toujours pas accès à internet. Certes, l'accès à ce moyen de communication s'est plutôt démocratisé ces dernières années, mais les personnes n'y ayant pas accès ont probablement des comportements bien différenciés des autres sur les variables d'intérêt. Collecter par internet pose ainsi un problème de couverture qui peut néanmoins être corrigé en partie, par exemple si la couverture peut s'expliquer largement par

les caractéristiques sociodémographiques.

L'autre gros problème des enquêtes par internet est celui des taux de réponse souvent très faibles. Ce problème de non-réponse ou de sélection est d'autant plus important qu'on peut supposer que les répondants ont des motivations intrinsèques qui différencient vraiment leurs réponses des autres y compris en contrôlant par des caractéristiques sociodémographiques.

Enfin, interroger par internet nécessite de gérer d'éventuels problèmes techniques. Dans la mesure du possible, la solution proposée doit être compatible avec les navigateurs web les plus utilisés. Cet élément est d'ailleurs plutôt en contradiction avec le fait d'offrir des questionnaires riches en contenu multimédia : avant de s'orienter dans cette direction, les instituts doivent proposer des questionnaires minimisant les possibilités de bugs pouvant détourner des répondants potentiels. Mais l'adaptation aux évolutions technologiques n'est pas aisée pour les instituts : alors que la question des designs optimaux de questionnaires internet sur des écrans d'ordinateur n'est pas encore réglée, se pose déjà la question des plateformes mobiles (smartphones, tablettes). Au Royaume-Uni, par exemple, près de 25 % des visites sur les sites internet gouvernementaux en 2013 s'effectuent avec des appareils mobiles (d'après le site du *Government Digital Service*). *A minima*, dans l'immédiat les questionnaires doivent être lisibles sur ces appareils ; une solution envisageable est que le questionnaire s'adapte au format et à la résolution de l'écran.

Pour faire face à ces différents problèmes, une solution peut être d'équiper une partie ou toute la population cible d'un dispositif permettant de répondre à l'enquête par internet (ordinateur simplifié, tablette) comme dans le cas des panels ELIPSS (France), GIP (Allemagne), LISS (Pays-Bas) ou Knowledge Network (États-Unis) (Blom *et al.*, 2015). Le panel ELIPSS est un panel représentatif de la population française où les enquêtés sont invités à répondre régulièrement à des enquêtes à finalité scientifique dans des domaines variés des sciences sociales. Au sein des logements sélectionnés, une personne est sollicitée pour participer au panel ELIPSS et se voit offrir à cette fin une tablette tactile ainsi qu'un abonnement internet 3G. (<http://www.elipss.fr/elipss/recruitment/about/>)

#### 2.4.6 Une revue de littérature

Ces dernières années, la littérature sur le multimode s'est enrichie, que cela soit par la recherche académique, ou par les expérimentations des instituts qui envisagent de généraliser cette manière de collecter l'information. L'un des principaux axes de recherche est celui des effets de mode, nous y consacrons le dernier chapitre de ce document. Les autres questions de recherche que nous allons brièvement présenter portent sur les taux de réponse, la manière de concevoir les différents questionnaires d'une enquête multimode et enfin la comparaison de la qualité des réponses obtenues dans les différents modes.

#### Taux de réponse et multimode impliquant internet

S'il existe des hiérarchies plus ou moins établies dans les différents modes de collecte concernant les taux de réponse, la combinaison de modes n'est pas toujours beaucoup plus efficiente



qu'une situation initiale où la collecte est unimode. Une question importante est celle de l'optimalité de la combinaison. Comment combiner les modes pour améliorer la réponse ? Que faire si l'on souhaite aussi diminuer les coûts ?

Millar et Dillman (2011) donnent ainsi un résultat qui peut paraître assez contre-intuitif. Offrir un choix simultané web et papier n'augmente pas significativement les taux de réponse, comparé à un protocole où le papier seul est proposé comme mode de réponse. En revanche, le multimode séquentiel, à savoir proposer d'abord le web puis, dans un second temps, le papier pour ceux qui n'ont pas répondu, donne de meilleurs résultats.

Holmberg, Lorenc et Werner (2010) regardent dans quelle mesure il est possible d'augmenter la proportion de réponses par internet lorsque plusieurs modes sont offerts. Ils montrent ainsi que pour un protocole de multimode séquentiel web puis papier, retarder l'envoi ou la mention du questionnaire papier a des conséquences limitées sur le taux de réponse global, mais joue énormément sur la proportion de réponses web.

## Questionnaires et multimode

Lorsque plusieurs modes de réponse sont utilisés pour collecter l'information auprès de différents sous-groupes de la population, à chaque mode de réponse doit correspondre un questionnaire. La question est alors de savoir dans quelle mesure ces différents questionnaires doivent être proches ou non. Il n'y a pas de réponse universelle car cela dépend des objectifs du protocole.

D'après de Leeuw et Dillman (2008), 4 approches sont possibles : la construction unimode, spécifique au mode, spécifique au mode préférentiel, et enfin l'approche généralisée.

Dans la construction unimode, on doit écrire et présenter les questions de la même manière quel que soit le mode. La construction spécifique profite des capacités spécifiques de chaque mode. Dans la construction spécifique au mode préférentiel, on privilégie la qualité des réponses sur un mode principal et l'équivalence entre les modes n'est pas recherchée à tout prix. Enfin, dans l'approche généralisée, les constructions des questionnaires sont différentes mais ceci avec un objectif d'équivalence du stimulus perçu pour obtenir des réponses équivalentes.

## Qualité des réponses sur internet

Internet étant un mode de réponse qui est amené à se développer et qui est déjà intégré dans les processus de collecte de certains instituts, une question que la recherche se pose est celle de la qualité des réponses obtenues par internet. En particulier, la recherche essaie de comparer ce mode de réponse à celui qui est à l'autre bout de l'échelle des coûts, le face à face.

L'hypothèse générale est que la tendance au *satisficing* est plus forte pour les questionnaires auto-administrés que pour les questionnaires dont la passation est effectuée par un enquêteur. Cela vaut en particulier lorsque l'on compare internet au face à face : l'étude empirique de Heerwegh et Loosveldt (2008) montre ainsi qu'il y a plus de non-réponses partielles sur internet ainsi qu'une moindre différenciation sur les échelles de réponse (sur une batterie de questions, le répondant internaute tend à choisir la même réponse et donc, la variance "*intra*" des réponses est moindre).

En théorie, un mode auto-administré comme internet est moins sujet aux biais de désirabilité sociale que ceux qui font intervenir un enquêteur ; par exemple, les individus hésitent moins en

auto-administré à choisir les modalités de réponses aux extrémités des échelles de réponse avec donc, pour une question isolée, une variance "*inter*" plus importante. Une étude d'Heerwegh (2009) montre ainsi que les biais de désirabilité sociale sont effectivement plus importants en face à face que sur internet mais il nuance en observant que ces biais sont un peu moins forts qu'attendus.

Ces derniers résultats, qui ne sont pas complètement généralisables car liés au thème, au questionnement ou encore au protocole de l'enquête analysée, constituent des exemples d'effets de mode que l'on peut souhaiter étudier sur des enquêtes spécifiques. Les aspects méthodologiques spécifiques à la question des effets de mode sont abordés dans le chapitre suivant.

## Chapitre 3

# L'ampleur des effets de mode

Une question importante lorsque l'on compare différents modes de collecte est de savoir si les réponses des enquêtés diffèrent ou non selon les modes. Ainsi, on définit les effets de mode comme les différences dans les réponses obtenues selon les modes de collecte. Cela peut porter sur les différences brutes entre les réponses, mais le plus souvent cela se réfère spécifiquement à l'erreur de mesure : toutes choses égales par ailleurs (donc à populations "comparables"), quelles sont les différences liées aux différents instruments de mesure ?

Il existe plusieurs manières de comparer les différents modes de collecte et la recherche sur le sujet s'enrichit régulièrement de résultats nouveaux. Néanmoins, on peut reprendre les propos de Couper (2011) qui affirme qu'"il n'y a pas encore de consensus sur l'étendue des différences d'erreurs de mesure entre les différents modes et sur ce (si besoin est) qu'il est nécessaire de faire à ce propos".

### 3.1 Comparaison naïve

Pour mesurer les effets de mode, une première idée est de comparer directement les résultats des différents modes de collecte.

Ainsi, considérons une variable d'intérêt  $Y$  et 2 types de questionnaires collectés<sup>1</sup> :

- cas  $M = A$  : des questionnaires recueillis via le mode de collecte  $A$
- cas  $M = B$  : des questionnaires recueillis via le mode de collecte  $B$

Une comparaison directe de  $\bar{Y}_{M=A}$  et  $\bar{Y}_{M=B}$  (ou tout autre paramètre de distribution) permet de mesurer l'ampleur des effets de mode au sens le plus global du terme mais ne permet pas d'évaluer ce qui nous intéresse le plus, à savoir les effets de mode liés à la mesure. Le souci de ce type de comparaison directe est que ce ne sont pas forcément les mêmes types de personnes qui répondent : les différences constatées peuvent être largement induites par la sélection des individus voire par des différences de couverture. Ce genre d'analyse ne permet pas ainsi de distinguer l'erreur de mesure, des erreurs de couverture ou de sélection, les différents effets étant confondus. Pour qu'une comparaison directe permette de conclure quant à l'effet de mode lié à

---

1. Ici, nous nous plaçons dans un cadre général où le processus de sélection des individus ou d'affectation aux modes de collecte (variable  $M$ ) est quelconque : il ne s'agit pas forcément d'une randomisation de type expérimentation contrôlée.

la mesure, les protocoles doivent être tels que les personnes sélectionnées soient les mêmes dans les différents modes : c'est ce qui est fait dans les expériences contrôlées dont nous verrons un exemple un peu plus loin.

## 3.2 Contrôle par les observables

Pour essayer de distinguer les effets de mode liés à la mesure, des effets de sélection et de couverture, une solution est de rendre les différents répondants "comparables" en contrôlant par des caractéristiques observables  $Z$ , pas ou très peu sensibles au mode de collecte. Ce type de contrôle est appelé *back-door model* dans certains papiers de la littérature sur les effets de mode (Vannieuwenhuyze *et al.*, 2014). On peut par exemple utiliser les techniques économétriques qui visent à évaluer l'effet d'un traitement, comme les modèles de régression ou les méthodes de matching. Une autre possibilité est de rendre comparables les populations par des techniques de repondération ou de calage.

Schématiquement, les différences brutes constatées dans les réponses constituent l'effet global de mode et les différences qui demeurent après un contrôle par des caractéristiques observables reflètent les effets de mode liés à la mesure, donc purgés des effets de sélection.

### 3.2.1 Exemples de contrôles par les observables

Les modèles de régression sont utilisés dans un article de Heerwelgh et Loosveldt pour comparer la qualité des réponses sur le web à celle des réponses en face à face (Heerwelgh et Loosveldt, 2008). Ainsi pour mesurer les effets du mode de collecte sur la non-réponse partielle et la différenciation sur l'échelle des réponses, les auteurs contrôlent de certaines caractéristiques sociodémographiques (âge, sexe, origine ethnique, etc.). Un contrôle sur variables sociodémographiques par une régression logistique est également réalisé dans l'article de Beck *et al.* pour comparer les estimations de prévalences de consommation d'alcool et de produits stupéfiants selon deux modes de collecte (téléphone et audio-CASI) (Beck *et al.*, 2014).

Un matching sur score de propension est mis en oeuvre par Ligtig *et al.* pour comparer les réponses obtenues par téléphone et par internet (Ligtig *et al.*, 2011). Le score de propension modélise la probabilité d'être un répondant via un certain mode plutôt que l'autre conditionnellement à des caractéristiques observables ( $P(M = A|Z)$ ). Ce score est ensuite utilisé pour appairer les individus des deux groupes de répondants aux scores proches et en déduire ensuite l'effet de mode purgé des effets de sélection. Un autre exemple d'utilisation du score de propension pour mesurer les effets de mode est réalisé par Thompson *et al.* sur les données de l'enquête *International Tobacco Control* (Thompson *et al.*, 2013).

Loosveldt et Sonck appliquent des méthodes de repondération (post-stratification et ajustement par score de propension) pour comparer les réponses d'une enquête en face à face en population générale avec celles d'un panel en ligne recruté par des méthodes non probabilistes (Loosveldt et Sonck, 2008).

### 3.2.2 Des effets toujours confondus ?

Dans ce genre d'analyse, l'hypothèse forte sous-jacente est que les caractéristiques observables suffisent pour expliquer entièrement les différences relevant de la sélection. Sous cette hypothèse, les différences dans les réponses conditionnellement aux caractéristiques observables (qui modélisent la sélection) sont les effets de mode liés à la mesure.

Le problème est que les variables de contrôle ne sont pas toujours suffisamment liées à la variable d'intérêt et que l'oubli d'une variable à la fois importante dans le processus de sélection et expliquant la variable d'intérêt peut potentiellement biaiser l'estimation des effets de mode. Là est toute la difficulté qui consiste à distinguer les effets de mode, des effets de sélection.

Typiquement, considérons que la "vraie" modélisation de la variable d'intérêt  $y$  est linéaire<sup>2</sup>

$$y = X_1\beta_1 + \alpha\mathbf{1}_{M=A} + \gamma Z + u$$

avec  $X_1$  un ensemble de caractéristiques observables,  $\mathbf{1}_{M=A}$  l'indicatrice de réponse selon le mode  $M = A$ ,  $Z$  une autre variable de sélection et  $u$  le terme d'erreur.

Cette modélisation constitue un cadre idéal où les variables aléatoires  $X_1$ ,  $Z$  et  $M$  (affectation ou choix du mode de collecte) sont orthogonales au terme d'erreur  $u$ . En particulier l'hypothèse  $M \perp u$  semble *a priori* très forte mais nous l'émettons dans l'objectif pédagogique de voir quelles peuvent être les conséquences de l'omission d'une variable fondamentale  $Z$  du processus de sélection sur l'estimation de l'effet de mode.

Dans le cas où les estimations sont faites sans inclure  $Z$  :

$$y = X_1\beta_1 + \alpha\mathbf{1}_{M=A} + v$$

Avec  $X = (X_1, \mathbf{1}_{M=A})$  et  $\beta = \begin{pmatrix} \beta_1 \\ \alpha \end{pmatrix}$ ,  $\hat{\beta}$  l'estimateur de  $\beta$  par le modèle complet,  $\beta^*$  par le modèle incomplet,

$$E(\hat{\beta}|X) = \beta$$

$$E(\beta^*|X) = \beta + (X'X)^{-1}X'Z\gamma$$

Lorsque nous avons plusieurs variables de contrôles, les prédictions sur les sens des biais potentiels ne sont pas aussi explicites que dans le cas d'une seule variable. Nous pouvons néanmoins dire que (Wooldridge, 2012) :

$$E(\alpha^*|X) = \alpha + \gamma\delta_Z^*$$

---

2. En particulier, ici nous faisons l'hypothèse restrictive que les effets des variables de contrôle ne sont pas différenciés suivant les modes de collecte : on pourrait relâcher cette hypothèse et, par exemple, étudier les différences sur variables d'intérêt à partir de décompositions de type Oaxaca-Blinder (Blinder, 1973; Oaxaca, 1973) par exemple sous la forme :

$$\bar{y}_{M=A} - \bar{y}_{M=b} = (\bar{X}_{M=A} - \bar{X}_{M=b})\beta_{M=A} + \bar{X}_{M=b}(\beta_{M=A} - \beta_{M=B})$$

avec  $\delta_Z^*$  le coefficient de la variable  $\mathbf{1}_{M=A}$  dans la régression auxiliaire de  $Z$  sur  $X = (X_1, \mathbf{1}_{M=A})$ .

Ainsi, en première approximation, on peut dire que l'ampleur du biais sur notre paramètre d'intérêt, censé mesurer les effets de mode liés à la mesure, risque d'être important quand :

- le lien entre la variable omise et la sélection est fort ;
- le pouvoir explicatif de la variable omise sur la variable d'intérêt est élevé.

Cette modélisation simplifiée montre que le choix des variables de contrôle est vraiment crucial lorsque l'on souhaite mesurer les effets de mode liés à la mesure. Des raisonnements similaires peuvent s'appliquer aux autres méthodes qui visent à rendre comparables les répondants selon les différents modes de collecte.

Le risque qui existe ainsi avec ce genre de méthode est de confondre des effets de sélection non contrôlés avec les effets de mode liés à la mesure. En effet, les individus qui répondent via un mode particulier (par exemple sur internet) gardent un profil spécifique que des variables de contrôle ne parviennent pas toujours à corriger (Frippiat et Marquis, 2010). Il n'est d'ailleurs pas toujours évident, même conceptuellement, de déterminer ce qui relève de la sélection ou de la mesure et il est sans doute illusoire de parvenir à une parfaite distinction. Par exemple, les résultats d'une expérimentation de l'Insee sur les conditions de logement (Amiel et Denoyelle, 2012) montrent qu'y compris après corrections sur caractéristiques sociodémographiques, les personnes qui répondent sur internet sont beaucoup plus pessimistes que celles qui répondent en face à face sur leurs conditions de logement. D'un côté, on peut interpréter ces différences comme un effet de mode lié à la mesure, principalement pour des questions de désirabilité sociale. De l'autre, on peut plutôt voir cela comme de la sélection non contrôlée avec par exemple un facteur inobservable de motivation ou d'intérêt pour le thème du questionnaire ayant un fort impact à la fois sur la sélection et sur la variable d'intérêt.

### 3.3 Expérimentation contrôlée

Pour évaluer les effets de mode liés à la mesure, une solution est l'expérimentation contrôlée : on affecte aléatoirement un mode de collecte, une fois que l'individu fait le choix de répondre à l'enquête. Dans ce cas là, il n'y a plus problème de sélection car, par construction, les différentes sous-populations de répondants sont identiques, du moins en espérance.

Le problème est qu'il est difficile de contraindre un individu à répondre dans un protocole d'enquête, or cette contrainte est théoriquement nécessaire pour ne pas avoir de problèmes liés à la non-réponse ou la sélection.

Un exemple d'expérimentation en laboratoire a été mené par Chang et Krosnick (2010). Leur objectif est de comparer des réponses auto-administrées sur ordinateur avec des réponses obtenues par entretien via un interphone. Une fois que les individus ayant accepté de participer à l'expérimentation arrivent au laboratoire, un mode de réponse leur est affecté aléatoirement. Les résultats de l'étude sont plutôt conformes à la théorie : les effets d'ordre des réponses, ici *recency*, à savoir la tendance à donner comme réponses celles proposées en dernier, sont plus marqués en entretien, les réponses sont moins socialement désirables sur ordinateur, etc.

Une telle expérimentation semble être une méthode particulièrement pertinente pour mesurer des effets de mode, mais se pose néanmoins la question du caractère extrapolable ou généralisable des résultats. En effet, non seulement les participants sont une population particulière puisqu'il s'agit d'étudiants en psychologie, mais l'incitation à participer n'est pas neutre non plus puisque cela leur permet de valider des crédits dans une matière.

### 3.4 Autres exemples de protocoles

D'autres protocoles ont été mis en œuvre ces dernières années pour essayer de mesurer les effets de mode liés à la mesure.

Vannieuwenhuyze *et al.* (2010) comparent le mode de collecte Capi (mode A) avec les modes de collecte Cati/Cawi (mode B) en exploitant 2 types d'enquêtes : une enquête multimode où les modes de réponse A et B sont possibles et une enquête où seul le mode de réponse A est proposé. L'expérimentation multimode que les auteurs exploitent a été réalisée en parallèle de la quatrième vague de l'enquête *European Social Survey* sur la base du même questionnaire : les variables comparées portent sur l'intérêt vis-à-vis de la politique, la complexité perçue de la politique et la participation électorale. Selon les auteurs, cette manière de procéder leur permet de bien distinguer les effets sur la mesure des effets de sélection. Leur méthode repose sur l'hypothèse, assez forte, dite de représentativité, selon laquelle les répondants à l'enquête effectuée via le mode A (enquête A) et les répondants à l'enquête multimode via les modes A et B (enquête A + B) représentent la même population. Selon les auteurs, cette hypothèse semble raisonnable lorsque tous les répondants de l'enquête A + B auraient également répondu à l'enquête A. L'hypothèse ne peut être complètement vérifiée mais les auteurs suggèrent une comparaison de caractéristiques sociodémographiques (peu sensibles au mode) des répondants pour donner quelques éléments *a minima* sur la pertinence de l'hypothèse. Dans leur discussion, les auteurs insistent néanmoins sur le fait que l'amplitude des biais de leurs estimations d'effets de mode (liés à la mesure) est d'autant plus importante que la proportion de personnes qui n'auraient répondu qu'à l'un des deux protocoles étudiés est forte. Notons par ailleurs qu'une telle hypothèse peut paraître légèrement contradictoire avec l'un des arguments avancés pour défendre le multimode, qui est que cette manière de collecter l'information permet d'atteindre des populations qui n'auraient pas répondu avec un seul mode proposé.

Les auteurs trouvent des effets de mode liés à la mesure qui relèvent essentiellement de la désirabilité sociale : les réponses Capi font état d'un plus grand intérêt à propos de la politique que les réponses Cati/Cawi et la politique est perçue comme moins complexe. En revanche, ils ne trouvent pas d'effet significatif sur la question de la participation électorale.

Buelens *et al.* (2012) mesurent des effets de mode en exploitant une enquête expérimentale qui fait intervenir 2 vagues d'interrogation : une première vague où différents modes sont assignés aléatoirement (web, papier, téléphone et face à face) et une deuxième vague où tout le monde répond avec le même mode (face à face). Les auteurs décomposent les effets de mode (mesurés relativement à un étalon qui est le face à face) en trois composantes : couverture, non-réponse et mesure. Pour cela, ils estiment des contrefactuels (configurations "virtuelles" où modes de sélection et mode de collecte différent), via des calages de type estimateurs par la régression généralisée (estimateurs GREG) à l'aide de variables auxiliaires recueillies à la

deuxième vague et de variables de source administrative. Une limite est que leur stratégie d'estimation repose sur une hypothèse assez forte, qui n'est pas testable : la non-réponse doit être ignorable pour les variables d'intérêt de la vague 1 conditionnellement aux variables de contrôle (variables recueillies à la vague 2 et variables de source administrative). Dans le cas où cette hypothèse n'est pas vérifiée, ce qui est considéré comme de l'erreur de mesure ou de couverture peut s'avérer être en réalité de l'erreur de non-réponse.

Le questionnaire de la vague 1 repose sur une version modifiée de l'enquête *Crime Victimization Survey* (CVS) à laquelle sont rajoutées des variables de l'enquête *Labour Force Survey* (LFS). Le questionnaire de la vague 2 fait intervenir certaines des variables clés de CVS, plusieurs questions attitudinales (politique, sécurité) ou encore des questions sur l'accès à internet et les préférences relatives aux différents modes de collecte. La conclusion des auteurs est que pour les variables LFS, l'effet de mode est un mélange d'effets liés à la couverture, la non-réponse et à la mesure et que l'effet de sélection peut être facilement corrigé à partir des variables de source administrative. En revanche, pour les variables CVS, l'effet de mode serait essentiellement de l'erreur de mesure.

### 3.5 Intérêt des expérimentations

Comme le dit la citation de Couper que nous avons évoquée plus haut, il n'existe pas vraiment à l'heure actuelle de consensus sur la manière de mesurer les effets de mode et la nature de leur ampleur. Par ailleurs, les différents résultats de la littérature sont difficilement généralisables car ils dépendent fortement du questionnaire et des thèmes abordés. Ces difficultés à généraliser rendent nécessaires les expérimentations dès lors que l'on souhaite modifier en profondeur un protocole d'enquête et introduire un nouveau mode de collecte.

Les protocoles expérimentaux permettent ainsi d'apporter des éléments de réponse quant à l'ampleur des effets de mode. Une première analyse des différences brutes entre les réponses selon les différents modes a déjà en soi un intérêt certain. La question est de voir ensuite si les différences persistent ou non une fois que l'on contrôle du mieux possible des effets de sélection et d'essayer d'interpréter au mieux les résultats obtenus.

Les expérimentations permettent également d'enrichir la connaissance quant à la pertinence de l'introduction ou non d'un nouveau mode de collecte dans une optique multimode. Quelle est la manière optimale d'introduire le nouveau mode de collecte ? Comment le mode doit être "coordonné" avec les autres modes de collecte déjà en place ? Quel est le rôle du nouveau mode de collecte, est-ce un mode majeur ou complémentaire ? Est ce que ce mode d'interrogation est "compatible" avec les objectifs principaux de l'enquête ?

Les expérimentations sont également un bon moyen d'accumuler de la connaissance sur des questions pratiques mais très importantes en termes de comparabilité de protocole qui tiennent par exemple à la sélection aléatoire du répondant au sein d'un ménage. Expérimenter n'est également pas une perte de temps dans la mesure où l'institut va acquérir de l'expérience lorsqu'il s'agit de tester de nouveaux modes de collecte qui ne sont pas encore généralisés dans les processus de production mais peut-être amenés à le devenir.



### 3.6 Corrections, ajustements

Une autre question d'importance où la recherche n'a pour le moment pas encore apporté de réponse claire porte sur ce qu'il est nécessaire de faire en termes de corrections et d'ajustements lorsqu'il existe des effets de mode. Est ce que l'on peut se satisfaire de la solution qui consiste à ne rien faire et donc à agréger les réponses quand elles proviennent de modes différents ?

Pour certains auteurs, la stratégie à mettre en oeuvre est essentiellement une stratégie de prévention : les questionnaires doivent être rédigés de manière à ce que les différences liées à la mesure soient minimales<sup>3</sup>. Mais d'autres auteurs préfèrent maximiser le potentiel de chacun des modes de collecte impliqués dans le protocole et lorsque des différences liées à la mesure existent, il faut se poser la question des ajustements à effectuer. Sur ce point, la littérature ne fournit pas encore de recommandations explicites<sup>4</sup>. Nous pouvons néanmoins cerner certaines des questions qu'il faut se poser. Sur la question de l'ajustement, il faut se demander par rapport à quoi l'on ajuste : est ce qu'un des modes peut être considéré comme le mode standard qu'il faut prendre en référence pour les ajustements ? Il n'est pas évident de répondre à cette question : la théorie peut nous donner des idées selon le type de questions, mais cela implique des postulats *a priori*, difficilement vérifiables, le concept de vraie valeur étant lui même difficile à appréhender. Une question que l'on doit se poser également porte sur la sensibilité des estimations que l'on fournit (en *output* de l'enquête donc après redressement, calage, correction des effets de sélection et de la non-réponse) à la répartition des réponses collectées entre les différents modes de collecte, que cette répartition résulte entièrement des choix des répondants ou qu'elle soit impliquée en partie par le protocole de l'enquête. Si ces estimations sont très sensibles à la répartition, sans doute faut-il se demander si la simple agrégation des réponses est vraiment la solution qui s'impose.

Une difficulté supplémentaire est qu'outre les problèmes éventuels liés aux différences selon les modes de collecte dans les fréquences, les niveaux des variables d'intérêt, il faut aussi se poser la question des relations entre variables : est-ce que les associations entre variables sont conservées (Charrance *et al.*, 2012) lorsque l'on bascule d'un mode à l'autre ou est-ce que des corrections sont également nécessaires ?

### 3.7 Quelques résultats sur les effets de mode

Pour les instituts de statistique, l'un des enjeux importants d'aujourd'hui est de développer la collecte multimode avec, en particulier, la possibilité pour les ménages de répondre sur internet. Mais comme nous l'avons vu dans ce document, cela pose de nombreuses questions méthodologiques. Comme les résultats de la recherche académique ne sont pas toujours généralisables, certains instituts de statistique expérimentent la collecte sur internet avant de l'introduire dans le processus de production de l'enquête concernée. Le principal objectif de ces expérimentations étant d'en savoir un peu plus sur les effets de mode.

La littérature insiste sur la dichotomie entre les modes où un enquêteur intervient et les modes où le questionnaire est auto-administré. Les réponses aux questionnaires administrés seraient

---

3. Approche unimode si l'on considère qu'une rédaction similaire des questionnaires minimise les différences de mesure ou approche généralisée (visant l'équivalence cognitive) si, au contraire, on pense que les spécificités des modes génèrent des différences de mesure lorsque les questionnaires sont rédigés à l'identique.

4. L'un des rares articles qui traite spécifiquement de ces questions est (Buelens et van den Brakel, 2014).

moins sujettes à la désirabilité sociale mais plus au *satisficing*. Typiquement, on constate que les répondants sont plus pessimistes (Lugtig *et al.*, 2011) en l'absence d'enquêteur quand on leur demande d'exprimer leurs opinions ou leurs attitudes. Mais il n'est pas toujours évident de savoir si les différences constatées après contrôle de la sélection sont des effets de mode liés à la mesure ou relèvent plutôt de la sélection non contrôlée.

### 3.7.1 Des répondants sur internet plus pessimistes

En 2010, une enquête expérimentale sur les conditions de logement a été réalisée par l'Insee (Amiel et Denoyelle, 2012) pour évaluer la qualité des données recueillies par internet en les comparant à celles issues de l'Enquête Nationale Logement 2006 (ENL 2006). Un premier constat est que le taux de réponse à l'enquête expérimentale est assez faible, de l'ordre de 20 %. Le contrôle des effets de sélection est effectué via un calage sur marges (âge, sexe, diplôme, revenu et diverses caractéristiques du logement) et après ce contrôle, des différences importantes persistent dans les réponses selon le mode de collecte (expérimentation vs ENL 2006). Compte tenu de l'importance de ces différences, on ne peut les imputer au décalage temporel existant entre les deux collectes : elles relèvent plutôt de l'effet de mode. Ainsi, les répondants sur internet sont beaucoup plus pessimistes sur de nombreux aspects de leurs conditions de logement : chauffage, sécurité, entretien, bruit, etc. Ces différences constatées entre les différents modes de collecte peuvent être liées à la mesure en tant que telle car la désirabilité sociale existe en présence d'un enquêteur. Mais il se peut également que les différences relèvent en partie d'effets de sélection non contrôlés : le taux de réponse n'étant pas très élevé sur internet, les répondants peuvent avoir un profil spécifique que le calage ne permet pas de contrôler (caractéristiques inobservables). Typiquement, on peut envisager que ceux qui répondent sur internet sont ceux qui ont particulièrement des choses à dire (et négatives) sur leurs conditions de logement.

### 3.7.2 Internet et papier : des différences ?

Les analyses des effets de mode ces dernières années portent le plus souvent sur la comparaison d'internet (questionnaire auto-administré) avec un mode où un enquêteur intervient, en face à face ou par téléphone. Mais la question de l'existence d'effets de mode entre deux modes auto-administrés mérite également d'être posée. Ainsi, les réponses internet et papier de l'enquête expérimentale Qualité de vie au travail 2013 ont été comparées sur certains aspects (Razafindranovona *et al.*, 2013). Deux protocoles ont été mis en oeuvre : une moitié de l'échantillon peut répondre soit par internet soit par papier<sup>5</sup> alors que pour l'autre moitié de l'échantillon, seule la réponse par internet est possible. Les taux de réponse sont de 50 % pour le protocole internet/papier (avec internet représentant environ les deux tiers de ces réponses collectées) et de 43 % pour le protocole internet seul.

Comme attendu, les répondants n'ont pas les mêmes caractéristiques : les répondants sur internet sont plus diplômés, sont plus souvent dans le haut de la distribution des revenus, etc. Pour évaluer les différences liées à la mesure, le contrôle des effets de sélection est effectué par une méthode économétrique de *kernel matching* sur score de propension (Heckman *et al.*, 1998).

---

5. Le questionnaire papier n'est envoyé qu'à la première relance alors que la réponse par internet est possible dès réception de la lettre-avis.

Les auteurs donnent deux principaux résultats relevant de l'effet de mode lié à la mesure qui portent sur :

- la tendance à utiliser ou non les différentes modalités sur les échelles de réponse, mesurée sur la base d'un indicateur de différenciation (McCarty et Shrum, 2000);
- le bien-être émotionnel des répondants, mesuré via un score appelé *WHO-5 index* développé par l'Organisation Mondiale de la Santé et imputé à partir des réponses à 5 questions de l'enquête.

Il s'avère que les répondants internet diffèrent plus leurs réponses (sur les échelles proposées) et que le bien-être émotionnel des répondants papier est plus élevé.

# Conclusion

Le recours à la collecte multimode avec possibilité de réponse par internet est promis à se développer dans les prochaines années pour enquêter les ménages ou les personnes. La collecte par internet est d'ailleurs d'ores déjà largement utilisée pour collecter l'information auprès des entreprises, la transition étant un peu moins problématique puisqu'internet a plutôt remplacé le papier, deux modes auto-administrés. Et même si les questions méthodologiques ne sont pas tout à fait de même nature, l'expérience accumulée dans le domaine de la collecte auprès des entreprises pourra être mise à profit pour converger vers des bonnes pratiques du questionnement en ligne. Au-delà des questions de coût, le recours à la collecte multimode ouvre de nombreuses perspectives. Cela pose aussi de nombreuses questions méthodologiques dont un cadre d'analyse pertinent est celui de l'erreur totale d'enquête avec sa décomposition selon les différentes sources d'erreur. Une autre question d'importance porte sur la place que doivent prendre les enquêtes par internet. Collecter l'information par des enquêtes par internet répond également à une nécessité d'adaptation des organisations enquêtrices au monde qui les entoure : internet, en tant que moyen de communication, est tellement présent dans le quotidien des personnes qu'il n'est guère concevable, dans le futur, de ne pas du tout y avoir recours pour collecter l'information. Reste que malgré tout, il s'agit d'un mode de collecte particulier, qui n'est pas du tout adapté pour faire passer en tant que telles, nombre d'enquêtes existantes, ayant toute leur légitimité et leur pertinence. Plus qu'un mode visant à remplacer définitivement la collecte en face à face ou par téléphone, internet est sans doute un mode plutôt amené à compléter (d'où la notion de multimode) d'une manière ou d'une autre les protocoles des enquêtes. Pour essayer de répondre aux questions méthodologiques qui se posent, et en particulier celles sur les effets de mode, un programme d'expérimentations d'enquêtes par internet (et papier) a été lancé à l'Insee. Ces expérimentations se déroulent en parallèle (ou avec un léger décalage pour éviter les interférences) des enquêtes principales (en face à face). Les résultats de ces expérimentations seront présentés dans des documents de travail à venir.

# Bibliographie

- ALEXANDER, M. G. et FISHER, T. D. (2003). Truth and consequences : Using the bogus pipeline to examine sex differences in self-reported sexuality. *Journal of sex research*, 40(1):27–35.
- AMIEL, M.-H. et DENOYELLE, T. (2012). Enquêtes en ligne : comparaison de modes de questionnement sur le thème du logement. *In Journées de Méthodologie Statistique*.
- ANDERSEN, R., KASPER, J. D., FRANKEL, M. R. et al. (1979). *Total survey error*. Jossey-Bass Publishers.
- ARDILLY, P. (2006). *Les techniques de sondage*. Editions Technip.
- BEAUMONT, J.-F. et HAZIZA, D. (2011). Theoretical framework for adaptive collection designs. *In International Total Survey Error Workshop*.
- BECK, F., GAUTIER, A., GUILBERT, P. et ARWIDSON, P. (2008). *Méthodes de sondages - Cours et cas pratiques*, chapitre L'acharnement téléphonique dans les enquêtes est-il justifié ?
- BECK, F., GUIGNARD, R. et LEGLEYE, S. (2014). Does Computer Survey Technology Improve Reports on Alcohol and Illicit Drug Use in the General Population ? A Comparison between Two Surveys with Different Data Collection Modes in France. *PLOS ONE*.
- BÉJIN, A. (1993). La masturbation féminine en france : Un exemple d'estimation et d'analyse de la sous-déclaration d'une pratique. *Population*, 48(5):1437–1450.
- BETHLEHEM, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4(3):251–260.
- BEUKENHORST, D. et WETZELS, W. (2009). A comparison of two mixed mode designs of the dutch safety monitor : Mode effects, costs, logistics. *In European Survey Research Association Conference*.
- BIEMER, P. P. (2010). *Latent class analysis of survey error*, volume 556. John Wiley & Sons.
- BIEMER, P. P. et LYBERG, L. E. (2003). *Introduction to survey quality*, volume 335. John Wiley & Sons.
- BLINDER, A. S. (1973). Wage discrimination : reduced form and structural estimates. *Journal of Human resources*, pages 436–455.
- BLOM, A., GATHMANN, C. et KRIEGER, U. (2015). Setting up an online panel representative of the general population : The german internet panel. *Field Methods*.
- BRENNER, M. (1982). Response effects of role-restricted characteristics of the interviewer. *Response behavior in the survey interview*, pages 131–165.
- BUELENS, B. et van den BRAKEL, J. A. (2014). Measurement error calibration in mixed-mode sample surveys. *Sociological Methods & Research*.

- BUELENS, B., van der LAAN, J., SCHOUTEN, B., van den BRAKEL, J., BURGER, J. et KLAUSCH, T. (2012). Disentangling mode-specific selection and measurement bias in social surveys. *Statistics Netherlands Discussion Papers*, 201211.
- CHANG, L. et KROSNICK, J. A. (2010). Comparing oral interviewing with self-administered computerized questionnaires: an experiment. *Public Opinion Quarterly*, 74(1):154–167.
- CHARRANCE, G., LEGLEYE, S. et RAZAFINDRATSIMA, N. (2012). Évaluation de la qualité d'utilisation d'Internet dans une enquête de santé publique en population générale. In *7ème colloque francophone sur les sondages*. SFdS.
- COUPER, M. (2008). *Designing effective web surveys*. Cambridge University Press Cambridge.
- COUPER, M. (2011). The future of modes of data collection. *Public opinion quarterly*, 75(5): 889–908.
- CROWNE, D. P. et MARLOWE, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, 24(4):349.
- DALENIUS, T. (1974). Ends and means of total survey design. *Forskningsprojektet Fel i Undersökningar*.
- DAVIS, R. E., COUPER, M., JANZ, N. K., CALDWELL, C. H. et RESNICOW, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, 25(1):14–26.
- de LEEUW, E. D. et DILLMAN, D. A. (2008). *International handbook of survey methodology*. Taylor & Francis US.
- de PERETTI, G. et RAZAFINDRANOVONA, T. (2013). Les enquêtes multimode : multi-solution ou multi-problème ? In *Journées de Statistique de la SFdS*.
- de PERETTI, G. et RAZAFINDRANOVONA, T. (2014). Les enquêtes multimode : attention aux effets de mode. *Statistique et société*, 2.
- DEMING, W. E. (1944). On errors in surveys. *American Sociological Review*, 9(4):359–369.
- DEMING, W. E. (1990). *Sample design in business research*, volume 23. Wiley. com.
- DESROSIÈRES, A. (2003). Les qualités des quantités. *Courrier des statistiques*, 105(106):51–63.
- DESROSIÈRES, A. (2012). Sur l'histoire de la méthodologie statistique : mesurer ou instituer ? deux traditions de recherche encore largement séparées. In *Congrès de la Société française de statistique*. SFdS.
- DILLMAN, D. (2009). *Mail and internet surveys : The tailored design method*, volume 2. Wiley New York.
- ECKMAN, S. (2012). Evidence of a coverage-nonresponse trade-off. In *International Total Survey Error Workshop*.
- EDWARDS, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37(2):90.
- FRIPPIAT, D. et MARQUIS, N. (2010). Les enquêtes par internet en sciences sociales : un état des lieux. *Population*, 65(2):309–338.
- GAUTIER, A., BECK, F., BLINEAU, V. et MOUSQUES, J. (2011). *Pratiques et Méthodes de sondage*, chapitre Le recours à une enquête web pour augmenter la participation à une enquête téléphonique : exemple du Baromètre santé médecins généralistes 2009. Dunod.
- GOMBAULT, V. (2013). L'internet de plus en plus prisé, l'internaute de plus en plus mobile. *Insee Première*, (1452).

- GOMBAULT, V. et DUÉE, M. (2012). Un exemple d'enquête multimode à l'Insee : l'enquête TIC auprès des ménages. *In 7ème colloque francophone sur les sondages*. SFdS.
- GOUX, D. (2003). Une histoire de l'enquête emploi. *Économie et statistique*, 362(1):41–57.
- GROVES, R. M. (1989). *Survey errors and survey costs*, volume 536. Wiley.
- GROVES, R. M. et LYBERG, L. (2010). Total survey error : Past, present, and future. *Public Opinion Quarterly*, 74(5):849–879.
- GROVES, R. M. et PEYTCHEVA, E. (2008). The impact of nonresponse rates on nonresponse bias : A meta-analysis. *Public opinion quarterly*, 72(2):167–189.
- HAGENAARS, J. A. et HEINEN, T. G. (1982). Effects of role-independent interviewer characteristics on responses. *Response behaviour in the survey-interview*, pages 91–130.
- HANSEN, M. H., HURWITZ, W. N., MARKS, E. S. et MAULDIN, W. P. (1951). Response errors in surveys. *Journal of the American Statistical Association*, 46(254):147–190.
- HECKMAN, J. J., ICHIMURA, H. et TODD, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2):261–294.
- HEERWEGH, D. (2009). Mode differences between face-to-face and web surveys : an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1):111–121.
- HEERWEGH, D. et LOOSVELDT, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population differences in response quality. *Public Opinion Quarterly*, 72(5):836–846.
- HOLBROOK, A. L., GREEN, M. C. et KROSNICK, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires : Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67(1):79–125.
- HOLMBERG, A., LORENC, B. et WERNER, P. (2010). Contact strategies to improve participation via the web in a mixed-mode mail and web survey. *Journal of Official Statistics*, 26(3):465.
- JÄCKLE, A., ROBERTS, C. et LYNN, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1):3–20.
- KISH, L. (1965). *Survey Sampling*. New York : Wiley.
- KREUTER, F. (2013). *Improving surveys with paradata : analytic uses of process information*. Wiley.
- KREUTER, F., MÜLLER, G. et TRAPPMANN, M. (2010). Nonresponse and measurement error in employment research : making use of administrative data. *Public Opinion Quarterly*, 74(5):880–906.
- KROSNICK, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3):213–236.
- LAGRANGE, H. (1991). Le nombre de partenaires sexuels : les hommes en ont-ils plus que les femmes ? *Population*, pages 249–277.
- LANGIS, P. et GERMAIN, B. (2009). *La sexualité humaine*. De Boeck Supérieur.
- LENGLART, F. (2013). Redesigning Household surveys ? Multimode data collection at Insee : challenges and opportunities. *In International Conference on Redesigning Household Surveys - Sharing experiences, Statistics Netherlands*.

- LESSLER, J. et KALSBECK, W. (1992). Nonsampling error in surveys.
- LITTLE, R. J. et RUBIN, D. B. (2002). Statistical analysis with missing data.
- LOOSVELDT, G. et SONCK, N. (2008). An evaluation of the weighting procedures for an online access panel survey. *In Survey Research Methods*, volume 2, pages 93–105. Wiley.
- LUGTIG, P., LENSVELT-MULDERS, G. J., FRERICHS, R. et GREVEN, F. (2011). Estimating nonresponse bias and mode effects in a mixed mode survey. *International Journal of Market Research*, 53(5):669–686.
- LYBERG, L. (2012). La qualité des enquêtes. *Techniques d'enquête*, 38(2):115–142.
- LYNN, P. (2008). The problem of non-response. *International Handbook of Survey Research Methods*.
- MCCARTY, J. A. et SHRUM, L. (2000). The measurement of personal values in survey research : A test of alternative rating procedures. *Public Opinion Quarterly*, 64(3):271–298.
- M McNABB, D. E. (2014). *Non Sampling Error in Social Surveys*. SAGE.
- MILLAR, M., O'NEILL, A. et DILLMAN, D. (2009). Are mode preferences real? Rapport technique, SERC.
- MILLAR, M. M. et DILLMAN, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, 75(2):249–269.
- OAXACA, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 14(3):693–709.
- OLSON, K. (2008). When do nonresponse follow-ups improve or reduce data quality? a meta-analysis of the existing literature. *In International Total Survey Error Workshop*.
- PAULHUS, D. L. (1991). Measurement and control of response bias.
- PEYTCHEV, A., RILEY, S., ROSEN, J., MURPHY, J. et LINDBLAD, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *In Survey Research Methods*, volume 4, pages 21–29.
- PLATEK, R. et SÄRNDAL, C.-E. (2001). Can a statistician deliver? *Journal of Official Statistics*, 17(1):1–20.
- RAZAFINDRANOVONA, T., de PERETTI, G. et BARRAU, A. (2013). The philosophy of french experiments on internet and mixed-mode data collection. *In UNECE Seminar on statistical data collection*.
- ROBERTS, C. (2007). Mixing modes of data collection in surveys : A methodological review. Rapport technique, ESRC NCRM.
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.
- SÄRNDAL, C.-E., SWENSSON, B. et WRÉTMAN, J. H. (2003). *Model assisted survey sampling*. Springer.
- SAUTORY, O. (2013). *La représentativité en statistique*, chapitre La représentativité à l'Insee. INED Méthodes et savoirs.
- SCHOUTEN, B., CALINESCU, M., LUITEN, A. M. C. et LUITEN, A. (2011). Optimizing quality of response through adaptive survey designs. Rapport technique, Statistics Netherlands.
- SCHOUTEN, B. et van DAM, G. (2013). Mixed-mode data collection - recent developments at statistics netherlands. *In UNECE Seminar on statistical data collection*.



- SIMON, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2):129.
- SIMON, H. A. (1982). *Models of bounded rationality : Empirically grounded economic reason*, volume 3. MIT press.
- THOMPSON, M. E., HUANG, Y. C., BOUDREAU, C., FONG, G. T., VAN DEN PUTTE, B., NAGELHOUT, G. E. et WILLEMSSEN, M. C. (2013). Mesurer les effets de la méthode de collecte des données. application à l'enquête international tobacco control. *Population*, 68(3).
- TOURANGEAU, R. (2000). *The psychology of survey response*. Cambridge University Press.
- TUCKEL, P. et O'NEILL, H. (2002). The vanishing respondent in telephone surveys. *Journal of Advertising Research*, 42(5):26–48.
- VANNIEUWENHUYZE, J., LOOSVELDT, G. et MOLENBERGHS, G. (2010). A method for evaluating mode effects in mixed-mode surveys. *Public opinion quarterly*, 74(5):1027–1045.
- VANNIEUWENHUYZE, J., LOOSVELDT, G. et MOLENBERGHS, G. (2014). Evaluating mode effects in mixed-mode survey data using covariate adjustment models. *Journal of Official Statistics*, 30(1):1–21.
- VERGER, D. (2013). Les enquêtes ménages par internet : les expériences récentes à l'insee. In *Séminaire de Méthodologie Statistique*.
- WATSON, J. (2009). Literature review identifying potential sources of mode effect between internet questionnaires and paper questionnaires. Rapport technique, ONS.
- WEISBERG, H. F. (2009). *The total survey error approach : A guide to the new science of survey research*. University of Chicago Press.
- WOOLDRIDGE, J. M. (2012). *Introductory econometrics : a modern approach*. Cengage Learning.
- YAN, T., RASINSKI, K., O' MUIRCHARTAIGH, C., KELLY, J., CAGNEY, P., JESSOE, R., et EULER, G. (2008). The dual task of interviewers. In *International Total Survey Error Workshop*.

## Série des Documents de Travail « Méthodologie Statistique »

**9601 :** Une méthode synthétique, robuste et efficace pour réaliser des estimations locales de population.  
**G. DECAUDIN, J.-C. LABAT**

**9602 :** Estimation de la précision d'un solde dans les enquêtes de conjoncture auprès des entreprises.  
**N. CARON, P. RAVALET, O. SAUTORY**

**9603 :** La procédure **FREQ** de **SAS** - Tests d'indépendance et mesures d'association dans un tableau de contingence.  
**J. CONFAS, Y. GRELET, M. LE GUEN**

**9604 :** Les principales techniques de correction de la non-réponse et les modèles associés.  
**N. CARON**

**9605 :** L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration.  
**P. RAVALET**

**9606 :** L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (**LOGIT**, **PROBIT**).  
**S. LOLLIVIER, M. MARPSAT, D. VERGER**

**9607 :** Enquêtes régionales sur les déplacements des ménages : l'expérience de Rhône-Alpes.  
**N. CARON, D. LE BLANC**

**9701 :** Une bonne petite enquête vaut-elle mieux qu'un mauvais recensement ?  
**J.-C. DEVILLE**

**9702 :** Modèles univariés et modèles de durée sur données individuelles.  
**S. LOLLIVIER**

**9703 :** Comparaison de deux estimateurs par le ratio stratifiés et application

aux enquêtes auprès des entreprises.

**N. CARON, J.-C. DEVILLE**

**9704 :** La faisabilité d'une enquête auprès des ménages.  
1. au mois d'août.  
2. à un rythme hebdomadaire

**C. LAGARENNE, C. THIESSET**

**9705 :** Méthodologie de l'enquête sur les déplacements dans l'agglomération toulousaine.  
**P. GIRARD.**

**9801 :** Les logiciels de désaisonnalisation **TRAMO** & **SEATS** : philosophie, principes et mise en œuvre sous **SAS**.  
**K. ATTAL-TOUBERT, D. LADIRAY**

**9802 :** Estimation de variance pour des statistiques complexes : technique des résidus et de linéarisation.  
**J.-C. DEVILLE**

**9803 :** Pour essayer d'en finir avec l'individu Kish.  
**J.-C. DEVILLE**

**9804 :** Une nouvelle (encore une !) méthode de tirage à probabilités inégales.  
**J.-C. DEVILLE**

**9805 :** Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish.  
**J.-C. DEVILLE**

**9806 :** Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel **POULPE**.  
**N. CARON, J.-C. DEVILLE, O. SAUTORY**

**9807 :** Estimation de données régionales à l'aide de techniques d'analyse multidimensionnelle.  
**K. ATTAL-TOUBERT, O. SAUTORY**

**9808 :** Matrices de mobilité et calcul de la précision associée.  
**N. CARON, C. CHAMBAZ**

**9809 :** Échantillonnage et stratification : une étude empirique des gains de précision.  
**J. LE GUENNEC**

**9810 :** Le Kish : les problèmes de réalisation du tirage et de son extrapolation.

**C. BERTHIER, N. CARON, B. NEROS**

**9901 :** Perte de précision liée au tirage d'un ou plusieurs individus Kish.  
**N. CARON**

**9902 :** Estimation de variance en présence de données imputées : un exemple à partir de l'enquête Panel Européen.  
**N. CARON**

**0001 :** L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (**LOGIT**, **PROBIT**) (version actualisée).  
**S. LOLLIVIER, M. MARPSAT, D. VERGER**

**0002 :** Modèles structurels et variables explicatives endogènes.  
**J.-M. ROBIN**

**0003 :** L'enquête 1997-1998 sur le devenir des personnes sorties du RMI - Une présentation de son déroulement.  
**D. ENEAU, D. GUILLEMOT**

**0004 :** Plus d'amis, plus proches ? Essai de comparaison de deux enquêtes peu comparables.  
**O. GODECHOT**

**0005 :** Estimation dans les enquêtes répétées : application à l'Enquête Emploi en Continu.  
**N. CARON, P. RAVALET**

**0006 :** Non-parametric approach to the cost-of-living index.  
**F. MAGNIEN, J. POUGNARD**

**0101 :** Diverses macros **SAS** : Analyse exploratoire des données, Analyse des séries temporelles.  
**D. LADIRAY**

**0102 :** Économétrie linéaire des panels : une introduction.  
**T. MAGNAC**

**0201 :** Application des méthodes de calages à l'enquête EAE-Commerce.  
**N. CARON**

**C 0201 :** Comportement face au risque et à l'avenir et accumulation patrimoniale - Bilan d'une expérimentation.  
**L. ARRONDEL, A. MASSON, D. VERGER**

**C 0202 :** Enquête Méthodologique Information et Vie Quotidienne - Tome 1 : bilan du test 1, novembre 2002.

**J.-A. VALLET, G. BONNET, J.-C. EMIN, J. LEVASSEUR, T. ROCHER, P. VRIGNAUD, X. D'HAULTFOEUILLE, F. MURAT, D. VERGER, P. ZAMORA**

**0203 :** General principles for data editing in business surveys and how to optimise it.  
**P. RIVIERE**

**0301 :** Les modèles logit polytomiques non ordonnés : théories et applications.  
**C. AFSA ESSAFI**

**0401 :** Enquête sur le patrimoine des ménages - Synthèse des entretiens monographiques.  
**V. COHEN, C. DEMMER**

**0402 :** La macro **SAS CUBE** d'échantillonnage équilibré  
**S. ROUSSEAU, F. TARDIEU**

**0501 :** Correction de la non-réponse et calage de l'enquête Santé 2002  
**N. CARON, S. ROUSSEAU**

**0502** : Correction de la non-réponse par ré pondération et par imputation  
**N. CARON**

**0503** : Introduction à la pratique des indices statistiques - notes de cours  
**J-P BERTHIER**

**0601** : La difficile mesure des pratiques dans le domaine du sport et de la culture - bilan d'une opération méthodologique  
**C. LANDRE, D. VERGER**

**0801** : Rapport du groupe de réflexion sur la qualité des enquêtes auprès des ménages  
**D. VERGER**

**M2013/01** : La régression quantile en pratique  
**P. GIVORD, X. D'HAULTFOEUILLE**

**M2014/01** : La microsimulation dynamique : principes généraux et exemples en langage R  
**D. BLANCHET**

**M2015/01** : la collecte multimode et le paradigme de l'erreur d'enquête totale  
**T. RAZAFINDROVONA**

**M2015/02** : Les méthodes de Pseudo-Panel  
**M. GUILLERM**