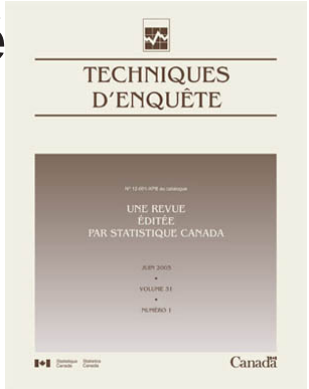
Composante du produit nº 12-001-X
Division des méthodes d'enquêtes auprès des entreprises

Article

Indicateurs de la représentativité de la réponse aux enquêtes



par Barry Schouten, Fannie Cobben et Jelke Bethlehem

Juin 2009

Statistique Canada Statistics Canada

Canada

Techniques d'enquête, juin 2009
Vol. 35, N° 1, pp. 107-121
Statistique Canada, N° 12-001-X au catalogue

Indicateurs de la représentativité de la réponse aux enquêtes

Barry Schouten, Fannie Cobben et Jelke Bethlehem [1]

Résumé

De nombreux organismes statistiques considèrent le taux de réponse comme étant l'indicateur de la qualité à utiliser en ce qui concerne l'effet du biais de non-réponse. Ils prennent donc diverses mesures en vue de réduire la non-réponse ou de maintenir la réponse à un niveau jugé acceptable. Cependant, à lui seul, le taux de réponse n'est pas un bon indicateur du biais de non-réponse. En général, un taux de réponse élevé n'implique pas que le biais dû à la non-réponse est faible. On trouve à cet égard de nombreux exemples dans la littérature (par exemple, Groves et Peytcheva 2006 ; Keeter, Miller, Kohut, Groves et Presser 2000 ; Schouten 2004).

Nous introduisons un certain nombre de concepts et un nouvel indicateur en vue d'évaluer la similarité entre la réponse à une enquête et l'échantillon de cette enquête. Cet indicateur de la qualité, que nous appelons indicateur R, peut servir de complément aux taux de réponse et est destiné principalement à évaluer le biais de non-réponse. Il peut faciliter l'analyse de la réponse aux enquêtes en fonction du temps, ou pour diverses stratégies d'enquête sur le terrain ou divers modes de collecte des données. Nous appliquons l'indicateur R à deux exemples pratiques.

Mots clés : Qualité ; non-réponse ; réduction de la non-réponse ; correction de la non-réponse.

1. Introduction

Il est bien décrit dans la littérature sur la méthodologie d'enquête que les taux de réponse sont, à eux seuls, de médiocres indicateurs du biais de non-réponse ; voir, par exemple, Curtin, Presser et Singer (2000), Groves, Presser et Dipko (2004), Groves (2006), Groves et Peytcheva (2006), Keeter et coll. (2000), Merkle et Edelman (2002), Heerwegh, Abts et Loosveldt (2007), ainsi que Schouten (2004). Cependant, les spécialistes du domaine n'ont pas encore établi d'autres indicateurs de la non-réponse qui seraient moins ambigus en tant qu'indicateurs de la qualité d'une enquête.

Nous proposons un indicateur, que nous appelons indicateur R (« R » pour représentativité), de la similarité entre la réponse à une enquête, d'une part, et l'échantillon ou la population à l'étude, d'autre part. Cette similarité peut être appelée « réponse représentative ». La littérature spécialisée offre de nombreuses interprétations du concept de « représentativité ». Consulter Kruskal et Mosteller (1979a, b et c) pour une étude approfondie de la littérature statistique et non statistique. Rubin (1976) a introduit le concept de non-réponse ignorable, c'est-à-dire les conditions minimales permettant d'obtenir une estimation sans biais d'une statistique. Certains auteurs définissent explicitement la représentativité. Hájek (1981) relie le terme « représentatif » à l'estimation de paramètres de population ; la paire formée par un estimateur et un mécanisme de création de données manquantes est représentative si, avec une probabilité de un, l'estimateur est égal au paramètre de population. Suivant la définition de Hajèk, les estimateurs par calage (par exemple, Särndal, Swensson et Wretman 2003) sont représentatifs pour les variables auxiliaires qui sont calées. Bertino (2006) définit un indice de représentativité dit univarié pour des variables aléatoires continues. Cet indice est une mesure exempte de distribution basée sur la statistique de Cramér-Von Mises. Kohler (2007) définit ce qu'il appelle un critère interne de représentativité. Ce critère univarié ressemble à la statistique Z pour les moyennes de population.

Nous isolons le concept de représentativité de l'estimation d'un paramètre de population particulier, mais nous le relions à l'effet sur la composition globale de la réponse. Dissocier les indicateurs d'un paramètre particulier permet de les utiliser comme outils pour comparer diverses enquêtes entre elles ou les résultats d'une enquête au cours du temps, ainsi que pour comparer des stratégies et modes différents de collecte des données. En outre, la mesure donne une perspective multidimensionnelle de la dissemblance entre l'échantillon et la réponse.

L'indicateur R que nous proposons s'appuie sur des probabilités de réponse estimées. L'estimation de ces probabilités implique que l'indicateur R proprement dit est une variable aléatoire et, par conséquent, qu'il possède une précision et éventuellement un biais. La taille de l'échantillon d'une enquête joue donc un rôle important dans l'évaluation de l'indicateur R, comme nous le montrerons. Cependant, cette dépendance existe pour toute mesure ; les petites enquêtes ne permettent tout simplement pas de tirer des conclusions catégoriques au sujet du mécanisme de création des données manquantes.

Nous montrons que l'indicateur R proposé est apparenté à la mesure V de Cramèr pour l'association entre la réponse

1. Barry Schouten, Fannie Cobben et Jelke Bethlehem, Statistics Netherlands, Department of Methodology and Quality, PO Box 4000, 2370 JM Voorburg, The Netherlands. Courriel : bstn@cbs.nl.

et les variables auxiliaires. En fait, nous considérons l'indicateur R comme une mesure du manque d'association. La situation est d'autant meilleure que l'association est faible, car cela signifie qu'il n'existe aucune preuve que la non-réponse a affecté la composition des données observées.

Afin de pouvoir utiliser les indicateurs R comme outils de surveillance et de comparaison de la qualité des enquêtes dans l'avenir, ceux-ci doivent avoir les caractéristiques d'une mesure. Autrement dit, nous voulons un indicateur R qui peut être interprété, mesuré et normalisé tout en ayant les propriétés mathématiques d'une mesure. Et ce, surtout parce que l'interprétation et la normalisation ne sont pas des caractéristiques directes.

Nous appliquons l'indicateur R à deux études menées à Statistiques Pays-Bas en 2005 et en 2006. Elles avaient pour objectifs de comparer différentes stratégies de collecte des données. Elles comportaient divers modes de collecte des données et diverses stratégies de suivi des cas de non-réponse. Chacune de ces études a été suivie d'une analyse détaillée des données et de la production de rapports. Par conséquent, elles conviennent bien pour une validation empirique de l'indicateur R. Nous comparons les valeurs de l'indicateur R aux conclusions des analyses. Nous renvoyons le lecteur à Schouten et Cobben (2007), ainsi qu'à Cobben et Schouten (2007) pour d'autres exemples et études empiriques.

À la section 2, nous commençons par discuter du concept de réponse représentative. Puis, à la section 3, nous définissons la notation mathématique pour notre indicateur. La section 4 est consacrée aux caractéristiques de l'indicateur R. La section 5 décrit l'application de cet indicateur R aux études sur le terrain. Enfin, à la section 6, nous présentons une discussion.

2. Le concept de réponse représentative

En premier lieu, nous expliquons ce que nous entendons quand nous disons qu'un ensemble de répondants à une enquête est représentatif de l'échantillon. Ensuite, nous rendons le concept de représentativité mathématiquement rigoureux en lui donnant une définition.

2.1 Que signifie représentatif ?

La littérature nous avertit de ne pas concentrer tous nos efforts sur les taux de réponse pour évaluer la qualité des enquêtes. Nous pouvons illustrer facilement ce point au moyen d'un exemple tiré de l'enquête néerlandaise POLS (acronyme de *Permanent Onderzoek Leefsituatie* ou Enquête intégrée sur les conditions de vie des ménages en français).

Le tableau 1 contient les estimations, fondées sur les réponses à l'enquête POLS après un mois et après deux mois, de la proportion de la population néerlandaise qui recevait une forme d'allocations sociales et de la proportion dont au moins un parent était né ailleurs qu'aux Pays-Bas. Pour les deux variables, les données sont extraites de registres et sont traitées artificiellement comme des réponses à une enquête en supprimant leurs valeurs pour les non-répondants. Les proportions dans l'échantillon sont également données au tableau 1. Après un mois, le taux de réponse était de 47,2 %, tandis qu'après la période complète d'interview de deux mois, il était de 59,7 %. Dans le cas de l'enquête POLS de 1998, le premier mois, la collecte a été effectuée par IPAO (interview sur place assistée par ordinateur). Après le premier mois, les non-répondants ont été affectés à un mode de collecte par ITAO (interview téléphonique assistée par ordinateur) s'ils étaient abonnés à un service téléphonique par fil dont le numéro était publié. Sinon, ils ont de nouveau été affectés à un mode de collecte par IPAO. Donc, le deuxième mois d'interview a donné 12,5 % supplémentaires de réponse. Cependant, l'examen du tableau 1 montre qu'après le deuxième mois, les estimations fondées sur les données de l'enquête présentent un biais plus important qu'après le premier mois.

Tableau 1
Moyennes des réponses à l'enquête POLS pour le premier mois de collecte et pour la période complète de collecte de deux mois

Variable	Après 1 mois	Après 2 mois	Échantillon
Recevant des allocations sociales	10,5 %	10,4 %	12,1 %
Parent non natif	12,9 %	12,5 %	15,0 %
Taux de réponse	47,2 %	59,7 %	100 %

L'exemple semble montrer clairement que l'effort accru a abouti à une réponse moins représentative en ce qui a trait aux deux variables auxiliaires. Mais d'une manière générale, qu'entendons-nous par représentatif ?

Le terme « représentatif » est souvent utilisé avec hésitation dans la littérature statistique. Kruskal et Mosteller (1979a, b et c) procèdent à un inventaire approfondi de l'utilisation du mot « représentatif » dans la littérature et relève neuf interprétations distinctes. Un certain nombre de ces interprétations sont omniprésentes dans les textes statistiques. Les interprétations statistiques que Kruskal et Mosteller ont nommées « absence de forces sélectives », « miniature de la population » et cas « typiques ou idéaux » se rapportent à l'échantillonnage probabiliste, l'échantillonnage par quota et l'échantillonnage raisonné. À la section suivante, nous proposons une définition de la représentativité qui correspond à l'« absence de forces sélectives ». Mais avant cela, nous expliquons pourquoi nous faisons ce choix.

Le concept de réponse représentative est aussi relié étroitement aux mécanismes de création des données manquantes désignés MCAR (pour *Missing-Completely-at-Random* ou Manquent entièrement au hasard), MAR (pour *Missing-at-Random* ou Manquent au hasard) et NMAR (pour *Not-Missing-at-Random* ou Ne manquent pas au hasard) qui sont souvent mentionnés dans la littérature ; voir Little et Rubin (2002). Un mécanisme de création de données manquantes est de type MCAR quand la probabilité de réponse ne dépend pas du sujet d'enquête d'intérêt. Le mécanisme est de type MAR si la probabilité de réponse dépend des données observées seulement, ce qui est donc une hypothèse plus faible que l'hypothèse MCAR. Si la probabilité dépend également des données manquantes, le mécanisme est de type NMAR. En fait, ces mécanismes ont pour origine la théorie statistique fondée sur la modélisation. Interprété plus ou moins librement en ce qui a trait à un sujet d'enquête, MCAR signifie que les répondants sont, en moyenne, les mêmes que les non-répondants, MAR signifie que, au sein de sous-populations connues, les répondants sont en moyenne les mêmes que les non-répondants, et NMAR implique que, même au sein de sous-populations, les répondants sont différents des non-répondants. L'ajout du sujet d'enquête est essentiel. Dans un même questionnaire, certains items peuvent être de type MCAR, tandis que d'autres sont de type MAR ou NMAR. En outre, l'hypothèse d'un mécanisme MAR pour un item tient pour une stratification particulière de la population. Un autre item pourrait nécessiter une stratification différente.

Étant donné que nous désirons surveiller et comparer la réponse à des enquêtes qui diffèrent par le sujet étudié ou par la période où elles sont exécutées, il n'est pas intéressant de définir une réponse représentative comme étant dépendante du sujet d'enquête proprement dit ni comme étant dépendante de l'estimateur utilisé. Nous préférons nous concentrer sur la qualité de la collecte des données plutôt que sur l'estimation. Ces conditions nous amènent à comparer la composition de la réponse (groupe de répondants) à celle de l'échantillon. Clairement, les sujets étudiés influencent la probabilité que les ménages participent à l'enquête, mais cette influence ne peut pas être mesurée ni testée et, par conséquent, de notre point de vue, ils ne peuvent pas représenter les données d'entrée utilisées pour évaluer la qualité de la réponse. Nous proposons de juger la composition de la réponse au moyen d'ensembles prédéfinis de variables qui sont observés en dehors du cadre de l'enquête et qui peuvent être employés pour chaque enquête examinée. Nous voulons que la sélection des répondants soit aussi proche que possible d'un « échantillon aléatoire simple de l'échantillon utilisé pour l'enquête », c'est-à-dire présentant une relation aussi faible que possible entre la réponse et les caractéristiques qui distinguent les unités les unes des autres. Le dernier énoncé peut être interprété comme signifiant l'absence de forces sélectives dans la sélection des répondants, ou comme un mécanisme de type MCAR en ce qui concerne toutes les variables étudiées possibles.

2.2 Définition d'un sous-ensemble de réponses représentatif

Soit $i = 1, 2, 3, \ldots, N$ les étiquettes d'unité pour la population. Nous désignons par s_i l'indicateur d'échantillon 0-1, c'est-à-dire que l'unité i prend la valeur 1 si elle est échantillonnée, et 0 autrement. Nous désignons par r_i l'indicateur de réponse 0-1 pour l'unité i. Si l'unité i est échantillonnée et qu'elle répond, alors $r_i = 1$. Sinon, sa valeur est 0. La taille d'échantillon est n. Enfin, π_i désigne la probabilité d'inclusion de premier ordre de l'unité i.

La clé de nos définitions réside dans les propensions à répondre individuelles. Soit ρ_i la probabilité que l'unité i réponde quand elle est échantillonnée.

L'interprétation d'une propension à répondre n'est pas directe. Nous suivons une approche assistée par modèle, ce qui signifie que seuls l'échantillon et les indicateurs de réponse ont un caractère aléatoire. Une probabilité de réponse est une caractéristique d'une unité étiquetée et identifiable, si l'on peut dire une pièce de monnaie biaisée que l'unité garde en poche, et est, par conséquent, inséparable de cette unité. Toutefois, moyennant un peu d'effort, tous les concepts peuvent être traduits dans un contexte fondé sur un modèle.

Pour commencer, nous donnons une définition forte.

Définition (*forte*) : *Un sous-ensemble de réponses est représentatif de l'échantillon si les propensions à répondre* ρ_i *sont les mêmes pour toutes les unités de la population*

$$\rho_i = P[r_i = 1 \mid s_i = 1] = \rho, \quad \forall i, \qquad (1)$$

et si la réponse d'une unité est indépendante de la réponse de toutes les autres unités.

Si un mécanisme de données manquantes satisfait la définition forte, il sera du type Manquent entièrement au hasard (MCAR) pour toutes les questions d'enquête possibles. Bien que la définition soit séduisante, il est impossible de tester sa validité en pratique. Nous ne possédons aucune répétition de la réponse d'une unité unique. Par conséquent, nous construisons aussi une définition faible qui peut être testée en pratique.

Définition (*faible*) : *Un sous-ensemble de réponses est représentatif d'une variable catégorique* X *possédant* H *catégories si la propension à répondre moyenne sur les catégories est constante, soit*

$$\bar{\rho}_h = \frac{1}{N_h}\sum_{k=1}^{N_h} \rho_{hk} = \rho, \quad \text{pour} \quad h = 1, 2, \ldots, H, \qquad (2)$$

où N_h est la taille de la population de la catégorie h, ρ_{hk} est la propension à répondre de l'unité k dans la classe h et la somme est faite sur toutes les unités dans cette catégorie.

La définition faible correspond à un mécanisme de création de données manquantes de type MCAR par rapport à X, car selon l'hypothèse MCAR, nous ne pouvons pas faire la distinction entre les répondants et les non-répondants en nous fondant sur la connaissance de X.

3. Indicateurs R

À la section précédente, nous avons défini une réponse fortement représentative et faiblement représentative. Les deux définitions s'appuient sur les probabilités de réponse individuelles qui sont inconnues en pratique. Pour commencer, nous prenons un indicateur R de population. Puis, nous basons le même indicateur R sur un échantillon et sur les propensions à répondre estimées.

3.1 Indicateurs R de population

Nous considérons d'abord la situation hypothétique où les propensions à répondre individuelles sont connues. Manifestement, dans ce cas, nous pouvons même tester la définition forte et nous cherchons simplement à mesurer la quantité de variation dans les propensions à répondre ; la représentativité au sens fort est d'autant moindre que la variation est forte. Soit $\rho = (\rho_1, \rho_2, \ldots, \rho_N)'$ un vecteur de propensions à répondre, soit $\mathbf{1} = (1, 1, \ldots, 1)'$ le vecteur de dimension N de valeurs 1, et soit $\rho_0 = \mathbf{1} \times \bar{\rho}$ le vecteur constitué de la propension à répondre moyenne de la population.

Toute fonction de distance d dans $[0, 1]^N$ suffirait à mesurer l'écart par rapport à une réponse fortement représentative en calculant $d(\rho, \rho_0)$. Notons que la hauteur de la réponse globale ne joue aucun rôle. La distance euclidienne est une fonction de distance simple. Appliquée à une distance entre ρ et ρ_0, cette mesure est proportionnelle à l'écart-type des probabilités de réponse

$$S(\rho) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (\rho_i - \bar{\rho})^2}. \qquad (3)$$

Il n'est pas difficile de montrer que

$$S(\rho) \le \sqrt{\bar{\rho}(1-\bar{\rho})} \le \frac{1}{2}. \qquad (4)$$

Nous voulons que l'indicateur R prenne les valeurs comprises dans l'intervalle [0, 1], la valeur 1 représentant une représentativité forte et la valeur 0 étant l'écart maximal par rapport à la représentativité forte. Nous proposons l'indicateur R, que nous désignons par R, défini par

$$R(\rho) = 1 - 2S(\rho). \qquad (5)$$

Notons que la valeur minimale de (5) dépend du taux de réponse (voir la figure 1). Pour $\bar{\rho} = 1/2$, il possède une valeur minimale de 0. Pour $\bar{\rho} = 0$ et $\bar{\rho} = 1$, aucune variation n'est manifestement possible et la valeur minimale est 1. Paradoxalement, la borne inférieure augmente quand le taux de réponse diminue pour passer de $1/2$ à 0. Dans le cas d'un taux de réponse faible, la possibilité que la variation des propensions à répondre individuelles soit forte est moindre.

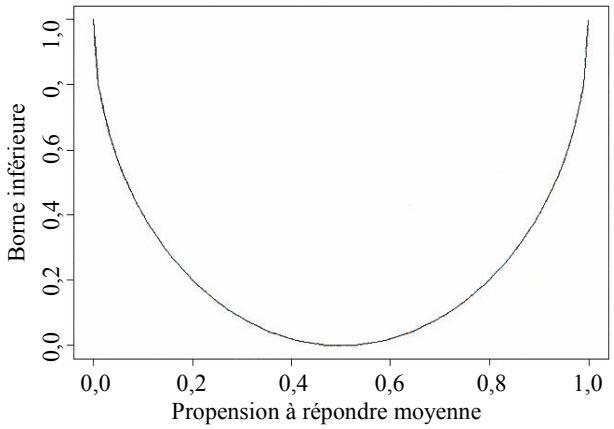


Figure 1 Valeur minimale de l'indicateur R (5) en fonction de la propension à répondre moyenne

Nous pouvons considérer R comme étant une mesure du manque d'association. Quand $R(\rho) = 1$, il n'existe aucune relation entre tout item d'une enquête et le mécanisme de création de données manquantes. Nous montrons que R est en fait étroitement lié à la statistique χ^2 bien connue qui est souvent utilisée pour tester l'indépendance et l'adéquation.

Supposons que les propensions à répondre diffèrent uniquement pour les classes h définies par une variable catégorique X. Soit $\bar{\rho}_h$ et f_h, respectivement, la propension à répondre et la fonction de population de la classe h, c'est-à-dire

$$f_h = \frac{N_h}{N}, \quad \text{pour} \quad h = 1, 2, \ldots, H. \qquad (6)$$

Donc, pour tout i avec $X_i = h$, la propension à répondre est $\rho_i = \bar{\rho}_h$.

Puisque la variance des propensions à répondre est la somme des variances « interclasses » et « intraclasse » sur

l'ensemble des classes h, et que nous supposons que les variances intraclasse sont nulles, il est vérifié que

$$S^2(\tilde{\rho}) = \frac{1}{N-1}\sum_{h=1}^{H} N_h(\bar{\rho}_h - \bar{\rho})^2$$

$$= \frac{N}{N-1}\sum_{h=1}^{H} f_h(\bar{\rho}_h - \bar{\rho})^2 \approx \sum_{h=1}^{H} f_h(\bar{\rho}_h - \bar{\rho})^2. \qquad (7)$$

La statistique χ^2 mesure la distance entre les proportions observées et prévues. Cependant, il ne s'agit d'une fonction de distance vraie qu'au sens mathématique pour les distributions marginales fixes f_h et $\bar{\rho}$. Nous pouvons appliquer la statistique χ^2 à X afin de « mesurer » la distance entre le comportement de réponse réel et le comportement de réponse qui est attendu quand la réponse est indépendante de X. Autrement dit, nous mesurons l'écart par rapport à la représentativité faible en ce qui concerne X.

Nous pouvons réécrire la statistique χ^2 pour obtenir

$$\chi^2 = \sum_{h=1}^{H} \frac{(N_h\bar{\rho}_h - N_h\bar{\rho})^2}{N_h\bar{\rho}}$$

$$+ \sum_{h=1}^{H} \frac{(N_h(1-\bar{\rho}_h) - N_h(1-\bar{\rho}))^2}{N_h(1-\bar{\rho})}$$

$$= \sum_{h=1}^{H} \frac{Nf_h(\bar{\rho}_h - \bar{\rho})^2}{\bar{\rho}} + \sum_{h=1}^{H} \frac{Nf_h(\bar{\rho}_h - \bar{\rho})^2}{(1-\bar{\rho})}$$

$$= \frac{N}{\bar{\rho}(1-\bar{\rho})}\sum_{h=1}^{H} f_h(\bar{\rho}_h - \bar{\rho})^2$$

$$= \frac{N-1}{\bar{\rho}(1-\bar{\rho})} S^2(\tilde{\rho}). \qquad (8)$$

Une mesure d'association qui converti la statistique χ^2 à l'intervalle [0, 1] (voir, par exemple, Agresti 2002) est le V de Cramèr

$$V = \sqrt{\frac{\chi^2}{N(\min\{C, R\} - 1)}}, \qquad (9)$$

où C et R sont, respectivement, le nombre de colonnes et de lignes dans la table de contingence sous-jacente. Le V de Cramèr atteint une valeur 0 si les proportions observées concordent exactement avec les proportions espérées et sa valeur maximale est 1. Dans notre cas, le dénominateur est égal à N puisque l'indicateur de réponse ne possède que deux catégories : réponse et non-réponse. Par conséquent, (9) se transforme en

$$V = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{N-1}{N\bar{\rho}(1-\bar{\rho})}}\, S(\tilde{\rho}). \qquad (10)$$

Partant de (10), nous voyons que, si N est grand, le V de Cramèr est approximativement égal à l'écart-type des propensions à répondre normalisées par l'écart-type maximal $\sqrt{\bar{\rho}(1-\bar{\rho})}$ pour une propension à répondre moyenne fixe $\bar{\rho}$.

3.2 Indicateurs R fondés sur la réponse

À la section 3.1, nous avons supposé que les propensions à répondre individuelles étaient connues. Naturellement, en pratique, elles ne le sont pas. En outre, dans un sondage, nous ne possédons des renseignements que sur le comportement de réponse des unités échantillonnées. Par conséquent, nous devons trouver d'autres solutions pour les indicateurs R. Un moyen évident consiste à utiliser des estimateurs fondés sur la réponse pour les propensions à répondre individuelles et la propension à répondre moyenne.

Soit $\hat{\rho}_i$ un estimateur de ρ_i qui utilise toutes les variables auxiliaires disponibles ou un sous-ensemble de celles-ci. Les méthodes qui permettent de calculer ce genre d'estimation sont, par exemple, les modèles de régression logit ou probit (Agresti 2002) et les arbres de classification CHAID (Kass 1980). Soit $\hat{\bar{\rho}}$ la moyenne pondérée d'échantillon des propensions à répondre estimées, c'est-à-dire

$$\hat{\bar{\rho}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\rho}_i \frac{s_i}{\pi_i}, \qquad (11)$$

où nous utilisons les poids d'inclusion.

Nous remplaçons R par l'estimateur $\hat{R}$

$$\hat{R}(\rho) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \frac{s_i}{\pi_i}(\hat{\rho}_i - \hat{\bar{\rho}})^2}. \qquad (12)$$

Notons que, dans (12), il existe en fait deux étapes d'estimation fondées sur des mécanismes probabilistes différents. Les propensions à répondre proprement dites sont estimées, ainsi que la variation dans les propensions. Nous revenons sur les conséquences de l'estimation en deux étapes à la section 4.

3.3 Exemple

Nous appliquons les indicateurs R proposés à des données provenant de l'enquête POLS de 1998 décrite à la section 2.1. Rappelons que l'enquête consistait en une combinaison d'interviews sur place et d'interviews téléphoniques dans laquelle les interviews du premier mois étaient de type IPAO uniquement. La taille de l'échantillon était de près de 40 000 et le taux de réponse était de 60 % environ. Nous avons relié des données administratives concernant le travail sur le terrain à l'échantillon et déduit si chaque tentative de prise de contact avait abouti à une réponse, ce

qui nous a permis de suivre la courbe de l'indicateur R durant la période de collecte des données.

Pour l'estimation des taux de réponse, nous nous sommes servis d'un modèle de régression logistique comprenant la région, les antécédents ethniques et l'âge comme variables indépendantes. La région était une classification à 16 catégories, c'est-à-dire les 12 provinces et les quatre villes les plus grandes (Amsterdam, Rotterdam, La Haye et Utrecht) comme catégories distinctes. Les antécédents ethniques étaient répartis en sept catégories : natif, Marocain, Turc, Surinamais, Néerlando-antillais, autre non-natif non occidental et autre non-natif occidental. La classification est fondée sur le pays de naissance des parents de la personne sélectionnée. La variable d'âge comprend trois catégories : de 0 à 34 ans, de 35 à 54 ans et de 55 ans et plus.

À la figure 2, $\hat{R}$ est représenté graphiquement en fonction du taux de réponse pour les six premières tentatives de prise de contact durant l'enquête POLS. La valeur la plus à gauche correspond à l'ensemble de répondants après une tentative. Pour chaque tentative supplémentaire, le taux de réponse augmente, mais l'indicateur révèle une diminution de la représentativité. Ce résultat confirme les constatations faites par Schouten (2004).

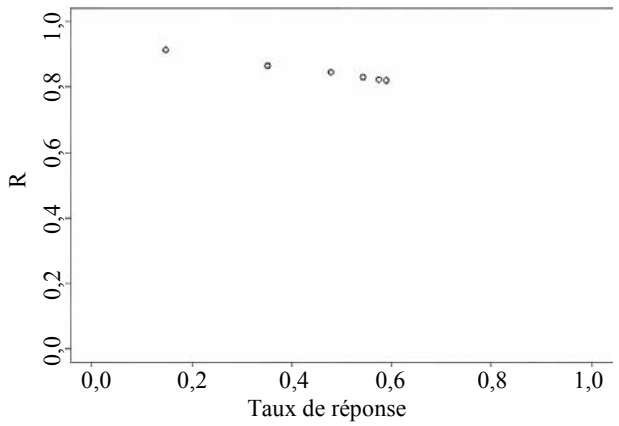


Figure 2 Indicateur R pour les six premières tentatives de prise de contact dans l'enquête POLS de 1998

4. Caractéristiques des indicateurs R

À la section 3, nous proposons un indicateur de la représentativité. Cependant, il est possible d'en construire d'autres. Les mesures d'association ou les indices d'adéquation sont nombreux, par exemple, Goodman et Kruskal (1979), Bentler (1990) et Marsh, Balla et McDonald (1988). La relation entre les mesures d'association et les indicateurs R est forte. Essentiellement, ces indicateurs ont pour objectif de mesurer, dans des conditions multivariées, le manque d'association. À la présente section, nous discutons des caractéristiques souhaitées des indicateurs R. Nous montrons que l'indicateur R proposé permet d'obtenir une borne supérieure simple pour le biais de non-réponse.

4.1 Caractéristiques générales

Nous voulons que les indicateurs R soient fondés sur une fonction ou une mesure de distance au sens mathématique. La propriété d'inégalité triangulaire d'une fonction de distance permet un classement partiel de la variation dans les propensions à répondre qui facilite l'interprétation. Une fonction de distance peut être dérivée facilement de toute norme mathématique. À la section 3, nous choisissons la norme euclidienne qui est celle utilisée couramment. Elle nous mène à un indicateur R qui s'appuie sur l'écart-type des propensions à répondre. D'autres normes, comme la norme supremum, ou norme sup, nous mèneraient à d'autres fonctions de distance. Toutefois, à la section 4.3, nous montrons que les indicateurs R fondés sur la norme euclidienne ont des caractéristiques de normalisation intéressantes.

Nous devons faire une distinction subtile entre les indicateurs R et les fonctions de distance. Ces dernières sont symétriques, tandis qu'un indicateur R mesure un écart par rapport à un point particulier, à savoir celui où toutes les propensions à répondre sont égales. Si nous modifions le vecteur des propensions individuelles, dans la plupart des cas, ce point se déplace. Toutefois, si nous fixons la valeur de la propension à répondre moyenne, la fonction de distance facilite l'interprétation.

À part une relation avec une fonction de distance, nous voulons pouvoir mesurer, interpréter et normaliser les indicateurs R. À la section 3.2, nous avons déjà dérivé des estimateurs fondés sur la réponse pour les indicateurs R « de population » qui ne sont pas mesurables quand les propensions à répondre sont inconnues et que l'on ne dispose que des réponses à un sondage. Donc, nous avons rendu les indicateurs R mesurables en passant à des estimateurs. Les deux autres caractéristiques sont examinées séparément dans les deux sections qui suivent.

4.2 Interprétation

La deuxième caractéristique des indicateurs R est la facilité avec laquelle nous pouvons interpréter leur valeur et le concept qu'ils mesurent. Nous sommes passés à un estimateur pour un indicateur R qui est fondé sur les échantillons des sondages et sur des estimateurs de probabilité de réponse individuelle. Ces deux éléments ont des conséquences très importantes en ce qui concerne l'interprétation et la comparaison des indicateurs R.

Puisque l'indicateur R est lui-même un estimateur, il est aussi une variable aléatoire. Autrement dit, il dépend de

l'échantillon, c'est-à-dire qu'il peut présenter un biais et qu'il possède une certaine exactitude. Mais qu'estime-t-il ?

Supposons d'abord que la taille d'échantillon est arbitrairement grande de sorte que la précision ne joue aucun rôle et que la sélection d'un modèle pour les propensions à répondre ne pose pas de problème. Autrement dit, nous sommes capables d'ajuster tout modèle à tout ensemble fixe de variables auxiliaires.

Il existe une forte relation entre l'indicateur R, d'une part, et la disponibilité et l'utilisation de variables auxiliaires, d'autre part. À la section 2, nous avons défini la représentativité forte et faible. Même dans le cas où nous arrivons à ajuster n'importe quel modèle, nous ne sommes pas capables d'estimer les propensions à répondre au-delà du « pouvoir de résolution » des variables auxiliaires disponibles. Donc, nous ne pouvons tirer des conclusions qu'au sujet de la représentativité faible par rapport à l'ensemble de variables auxiliaires. Cela veut dire que, chaque fois que nous utilisons un indicateur R, nous devons utiliser comme complément de sa valeur l'ensemble de covariables qui ont servi de grille pour estimer les propensions à répondre individuelles. Si l'indicateur R est utilisé pour des comparaisons, les ensembles de variables doivent être les mêmes. Il convient de souligner qu'il n'est pas nécessaire d'utiliser toutes les variables auxiliaires pour l'estimation des propensions à répondre, puisqu'elles peuvent ne pas accroître la puissance explicative du modèle. Cependant, des ensembles identiques devraient être disponibles. L'indicateur R mesure alors un écart par rapport à la représentativité faible.

L'indicateur R ne reflète pas les différences entre les probabilités de réponse à l'intérieur d'autres sous-groupes de la population que ceux définis par les classes de X. Si nous désignons de nouveau par $h = 1, 2, \ldots, H$ les strates définies par X, par N_h la taille de la strate h et par $\bar{\rho}_h$ la moyenne de population des probabilités de réponse dans la strate h, il n'est pas difficile de montrer que $\hat{R}$ est une estimateur convergent de

$$R_X(\rho) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{h=1}^{H} N_h(\bar{\rho}_h - \bar{\rho})^2}, \qquad (13)$$

quand des modèles standard, comme la régression logistique ou la régression linéaire, sont utilisés pour estimer les probabilités de réponse. Naturellement, les indicateurs (13) et (5) peuvent être différents.

En pratique, la taille d'échantillon n'est pas arbitrairement grande. Elle a une incidence sur les deux étapes d'estimation, c'est-à-dire l'estimation des propensions à répondre et celle de l'indicateur R en utilisant un échantillon.

Si nous connaissions les propensions à répondre individuelles, l'estimation fondée sur l'échantillon de l'indicateur R ne donnerait lieu qu'à une variance et ne produirait pas de biais. Nous serions capables d'estimer l'indicateur R de population sans biais. Donc, pour les petites tailles d'échantillon, les estimateurs auraient une faible précision dont il pourrait être tenu compte en utilisant des intervalles de confiance au lieu de simples estimateurs ponctuels.

Les incidences sur l'estimation des probabilités de réponse sont toutefois différentes, à cause de la sélection et de l'ajustement du modèle. Il existe deux options. Nous pouvons imposer un modèle pour estimer les propensions à répondre en fixant les covariables au préalable ou nous pouvons laisser le modèle dépendre des covariables dont la contribution est significative par rapport à un niveau prédéfini. Dans le premier cas, nous n'introduirons de nouveau aucun biais, mais l'erreur-type pourrait être affectée par un surajustement. Dans le deuxième cas, le modèle pour l'estimation des propensions à répondre dépend de la taille de l'échantillon ; plus l'échantillon est grand, plus le nombre d'interactions acceptées comme étant significatives est élevé. Même si l'ajustement de modèles en se fondant sur un niveau de signification est une pratique statistique courante, la sélection du modèle peut introduire un biais et une variance dans l'estimation de tout indicateur R. Ce point est facile à comprendre si l'on considère la situation extrême d'un échantillon de taille 10. Pour un si petit échantillon, aucune interaction entre le comportement de réponse et les caractéristiques auxiliaires ne sera acceptée, ce qui laisse un modèle vide et un indicateur R estimé de 1. Les petits échantillons ne permettent tout simplement pas d'estimer les propensions à répondre. En général, un échantillon de petite taille conduira donc à une vue plus optimiste de la représentativité.

Nous devons faire ici une autre distinction subtile. Il se pourrait que, dans un sondage, un grand nombre d'interactions réunies contribuent à la prédiction des propensions à répondre, mais que les contributions individuelles soient très faibles, tandis que dans un autre sondage, une seule interaction intervienne, mais que sa contribution soit grande. Il se peut qu'individuellement, aucune petite contribution ne soit significative, mais qu'ensemble, elles soient aussi fortes que la grande contribution unique qui est significative. Donc, nous serions plus optimistes dans le premier exemple, même pour des tailles d'échantillon comparables.

Ces observations montrent qu'il faut toujours utiliser les indicateurs R avec prudence. Ils ne peuvent pas être considérés indépendamment des variables auxiliaires qui ont été utilisées pour les calculer. En outre, la taille de l'échantillon a une incidence sur le biais ainsi que sur la précision.

4.3 Normalisation

La troisième caractéristique importante d'un indicateur R est la normalisation. Nous voulons pouvoir donner des bornes à un indicateur R afin que son échelle, et donc, les variations de sa valeur, aient une signification. De toute

évidence, les problèmes d'interprétation soulevés à la section précédente ont également une incidence sur la normalisation de l'indicateur R. Par conséquent, ici, nous supposons que nous nous trouvons dans la situation idéale où nous pouvons estimer les propensions à répondre sans biais. Cette hypothèse tient pour les grandes enquêtes. Nous discutons la normalisation de l'indicateur R correspondant à $\hat{R}$.

4.3.1 Biais absolu maximal et racine carrée de l'erreur quadratique moyenne

Nous montrons que, pour tout item d'une enquête Y, l'indicateur R peut être utilisé pour imposer des bornes supérieures au biais de non-réponse et à la racine carrée de l'erreur quadratique moyenne (REQM) des moyennes ajustées des réponses. Nous appliquons ces bornes à l'indicateur R pour montrer l'effet sous les pires scénarios.

Soit Y une variable qui est mesurée dans une enquête et soit $\hat{\bar{y}}_{\mathrm{HT}}$ l'estimateur Horvitz-Thompson de la moyenne de population basée sur les réponses à l'enquête. On peut montrer (par exemple, Bethlehem 1988, Särndal et Lundström 2005) que son biais $B(\hat{\bar{y}}_{\mathrm{HT}})$ est approximativement égal à

$$B(\hat{\bar{y}}_{\mathrm{HT}}) = \frac{C(y,\ \rho)}{\bar{\rho}}, \tag{14}$$

avec $C(y,\ \rho) = 1/N \sum_{i=1}^{N}(y_i - \bar{y})\,(\rho_i - \bar{\rho})$ la covariance de population entre les items étudiés et les probabilités de réponse. Pour une approximation étroite de la variance $s^2(\hat{\bar{y}}_{\mathrm{HT}})$ de $\hat{\bar{y}}_{\mathrm{HT}}$, nous nous référons à Bethlehem (1988).

L'inégalité de Cauchy-Schwarz permet de trouver une normalisation de R. Cette inégalité énonce que la covariance entre deux variables quelconques est bornée au sens absolu par le produit des écarts-types des deux variables. Nous pouvons traduire cela en bornes pour le biais (14) de $\hat{\bar{y}}_{\mathrm{HT}}$

$$|B(\hat{\bar{y}}_{\mathrm{HT}})| \le \frac{S(\rho)S(y)}{\bar{\rho}} = \frac{(1 - R(\rho))S(y)}{2\bar{\rho}} = B_m(\rho,\ y). \tag{15}$$

Manifestement, nous ne connaissons pas la borne supérieure $B_m(\rho,\ y)$ dans (15), mais nous pouvons l'estimer en utilisant l'échantillon et les probabilités de réponse estimées. Nous désignons l'estimateur par $\hat{B}_m(\hat{\rho},\ y)$.

De la même manière, nous pouvons établir une borne pour la racine carrée de l'erreur quadratique moyenne (REQM) de $\hat{\bar{y}}_{\mathrm{HT}}$. Il est vérifié approximativement que

$$\mathrm{REQM}(\hat{\bar{y}}_{\mathrm{HT}}) = \sqrt{B^2(\hat{\bar{y}}_{\mathrm{HT}}) + s^2(\hat{\bar{y}}_{\mathrm{HT}})} \le \sqrt{B_m^2(\rho,\ y) + s^2(\hat{\bar{y}}_{\mathrm{HT}})} = E_m(\rho,\ y). \tag{16}$$

De nouveau, nous ne connaissons pas $E_m(\rho,\ y)$. À sa place, nous utilisons l'estimateur fondé sur l'échantillon qui emploie les probabilités de réponse estimées, désignées par $\hat{E}_m(\hat{\rho},\ y)$.

Les bornes $\hat{B}_m(\hat{\rho},\ y)$ et $\hat{E}_m(\hat{\rho},\ y)$ sont différentes pour chaque item y de l'enquête. Pour les comparaisons, il est par conséquent commode de définir un item d'enquête hypothétique. Nous supposons que $\hat{S}(y) = 0{,}5$. Nous désignons les bornes correspondantes par $\hat{B}_m(\hat{\rho})$ et $\hat{E}_m(\hat{\rho})$. Elles sont égales à

$$\hat{B}_m(\hat{\rho}) = \frac{(1 - \hat{R}(\hat{\rho}))}{4\hat{\bar{\rho}}} \tag{17}$$

$$\hat{E}_m(\hat{\rho}) = \sqrt{\hat{B}_m^2(\hat{\rho}) + \hat{s}^2(\hat{\bar{y}}_{\mathrm{HT}})}. \tag{18}$$

Nous calculons (17) et (18) dans le cas de toutes les études décrites à la section 5. Nous devons souligner que (17) et (18) sont de nouveau des variables aléatoires qui possèdent une certaine précision et qui sont susceptibles de présenter un biais.

4.3.2 Fonctions de représentativité de la réponse

À la section précédente, nous avons utilisé l'indicateur R pour déterminer les bornes supérieures du biais de non-réponse et de la racine carrée de l'erreur quadratique moyenne de la moyenne (ajustée) des réponses. Inversement, nous pouvons établir une borne inférieure de l'indicateur R en demandant que le biais absolu de non-réponse ou la racine carrée de l'erreur quadratique moyenne soit inférieur à une valeur prescrite. Cette borne inférieure peut être choisie comme étant l'un des ingrédients des contraintes de qualité imposées aux données d'enquête par l'utilisateur de ces données. Si ce dernier ne veut pas que le biais de non-réponse ou la racine carrée de l'erreur quadratique moyenne excède une certaine valeur, l'indicateur R doit être supérieur à la borne correspondante.

Il est évident que les bornes inférieures de l'indicateur R dépendent de l'item d'enquête. Par conséquent, nous nous limitons de nouveau à un item hypothétique pour lequel $\hat{S}(y) = 0{,}5$.

Il est facile de montrer, en partant de (17), que si nous demandons que

$$\hat{B}_m(\hat{\rho}) \le \gamma, \tag{19}$$

il doit être vérifié que

$$\hat{R} \ge 1 - 4\hat{\bar{\rho}}\gamma = r_1(\gamma,\ \hat{\rho}). \tag{20}$$

De manière analogue, en utilisant (18) et en demandant que

$$\hat{E}_m(\hat{\rho}) \le \gamma, \tag{21}$$

nous arrivons à

$$\hat{R} \geq 1 - 4\hat{\bar{\rho}}\sqrt{\gamma^2 - \hat{s}^2(\hat{\bar{y}}_{\mathrm{HT}})} = r_2(\gamma, \hat{\rho}). \qquad (22)$$

Dans (20) et (22), nous représentons par $r_1(\gamma, \hat{\rho})$ et $r_2(\gamma, \hat{\rho})$ les bornes inférieures de l'indicateur R. Dans la section qui suit, nous donnons à $r_1(\gamma, \hat{\rho})$ et $r_2(\gamma, \hat{\rho})$ le nom de fonctions de représentativité de la réponse. Nous les calculons pour les études décrites à la section 5.

4.3.3 Exemple

Nous illustrons de nouveau la normalisation au moyen de l'exemple utilisé aux sections 2.1 et 3.3. La figure 3 contient la fonction de représentativité de la réponse $r_1(\gamma, \hat{\rho})$ et les indicateurs R observés $\hat{R}$ pour les six tentatives de prise de contact durant l'enquête POLS de 1998. Nous avons choisi trois valeurs de γ, soit $\gamma = 0{,}1$; $\gamma = 0{,}075$ et $\gamma = 0{,}05$.

La figure 3 indique qu'après la deuxième tentative de prise de contact, les valeurs de l'indicateur R sont supérieures à la borne inférieure correspondant au niveau de 10 %. Après quatre tentatives, l'indicateur R est proche du niveau de 7,5 %. Cependant, les valeurs ne sont jamais supérieures à l'autre borne inférieure qui est fondée sur le niveau de 5 %.

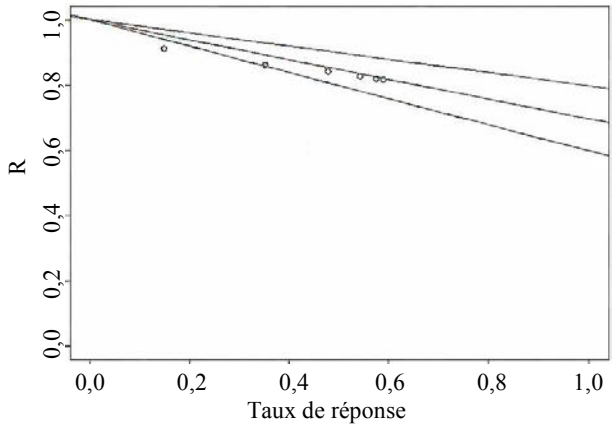


Figure 3 Bornes inférieures de l'indicateur R $\hat{R}$ pour les six premières tentatives de prise de contact durant l'enquête POLS de 1998. Les bornes inférieures sont basées sur $\gamma = 0{,}1$, $\gamma = 0{,}075$ et $\gamma = 0{,}05$

À la figure 4, le biais absolu maximal $\hat{B}_m(\hat{\rho})$ est représenté en fonction du taux de réponse pour les six tentatives de prise de contact. Après la troisième tentative, l'indicateur R a convergé vers une valeur autour de 8 %.

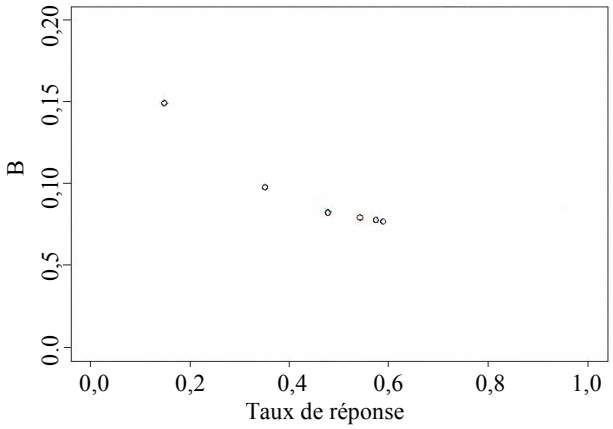


Figure 4 Biais absolu maximal pour les six premières tentatives de prise de contact dans l'enquête POLS de 1998

5. Application de l'indicateur R

À la présente section, nous appliquons l'indicateur R à des études portant sur diverses stratégies de suivi des cas de non-réponse et diverses combinaisons de modes de collecte des données. La première a trait à l'Enquête sur la population active (EPA) des Pays-Bas. Elle comprend un examen de l'approche du rappel (Hansen et Hurwitz 1946) et de celle des questions de bases (Kersten et Bethlehem 1984). La deuxième concerne les plans de collecte de données à mode mixte appliqués dans le cadre de l'enquête nationale de surveillance de la sécurité (*National Safety Monitor*) des Pays-Bas.

Aux sections 5.2 et 5.3, nous examinons ces études de plus près en ce qui a trait à la représentativité de leurs différentes stratégies de travail sur le terrain. Avant cela, à la section 5.1, nous décrivons comment nous obtenons une approximation des erreurs-types.

5.1 Erreur-type et intervalle de confiance

Si nous voulons comparer les valeurs de l'indicateur R pour diverses enquêtes ou stratégies de collecte des données, nous devons estimer leurs erreurs-types.

L'indicateur R $\hat{R}$ fait intervenir l'écart-type d'échantillon des probabilités de réponse estimées. Autrement dit, deux processus aléatoires entrent en jeu. Le premier est l'échantillonnage de la population, et le second, le mécanisme de réponse des unités échantillonnées. Si nous connaissions les probabilités de réponse réelles, le tirage d'un échantillon introduirait quand même une incertitude au sujet de l'indicateur R de population, et donc, entraînerait une certaine perte de précision. Comme nous ne connaissons pas les probabilités de réponse réelles, nous les

estimons en utilisant l'échantillon, ce qui cause une perte de précision supplémentaire.

Un calcul analytique de l'erreur-type de $\hat{R}$ n'est pas simple à cause de l'estimation des probabilités de réponse. Ici, nous nous résignons à utiliser des approximations numériques naïves de l'erreur-type. Nous estimons l'erreur-type de l'indicateur R par une méthode du bootstrap non paramétrique (Efron et Tibshirani 1993). Selon cette méthode, l'erreur-type de l'indicateur R est estimée en sélectionnant un nombre $b = 1, 2, \ldots, B$ d'échantillons bootstrap. Il s'agit d'échantillons tirés indépendamment et avec remise à partir de l'ensemble de données originales et de même taille n que cet ensemble. L'indicateur R est calculé pour chaque échantillon bootstrap b. Nous obtenons donc B répliques de l'indicateur R ; $\hat{R}_b^{\mathrm{BT}}$, $b = 1, 2, \ldots, B$. L'erreur-type pour la loi empirique de ces B répliques est une estimation de l'erreur-type de l'indicateur R, c'est-à-dire

$$s_R^{\mathrm{BT}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{R}_b^{\mathrm{BT}} - \hat{\bar{R}}^{\mathrm{BT}})^2} \qquad (23)$$

où $\hat{\bar{R}}^{\mathrm{BT}} = 1/B \sum_{b=1}^{B} \hat{R}_b^{\mathrm{BT}}$ est l'indicateur R moyen estimé.

Dans les approximations, nous prenons $B = 200$ pour toutes les études. Nous avons expérimenté de plus grandes valeurs de B allant jusqu'à $B = 500$, mais avons constaté dans tous les cas que l'estimation de l'erreur-type avait convergé dès la valeur $B = 200$.

Nous déterminons les intervalles de confiance $100(1 - \alpha)\,\%$ en supposant que la loi de $\hat{R}$ est approximativement normale et en employant les erreurs-types estimées données par (23), soit

$$\mathrm{CI}_\alpha^{\mathrm{BT}} = (\hat{R} \pm \xi_{1-\alpha} \times s_R^{\mathrm{BT}}) \qquad (24)$$

où $\xi_{1-\alpha}$ est le quantile $1 - \alpha$ de la loi normale standard.

5.2 Enquête sur la population active – Étude de suivi de 2005

De juillet à décembre 2005, Statistique Pays-Bas a réalisé un suivi à grande échelle des cas de non-réponse à l'Enquête sur la population active (EPA) des Pays-Bas. L'étude a consisté en une nouvelle tentative d'interview auprès de deux échantillons de non-répondants à l'EPA, selon l'approche du rappel (Hansen et Hurwitz 1946) dans un cas et selon l'approche des questions de base (Kersten et Bethlehem 1984) dans l'autre. Les échantillons étaient formés de ménages sélectionnés pour participer à l'EPA qui avaient refusé de participer, dont les réponses n'avaient pas été traitées ou avec lesquels il n'y avait pas eu prise de contact pour l'EPA de juillet à octobre. Pour concevoir l'étude de suivi, nous avons tenu compte des recommandations émanant des études de Stoop (2005) et de Voogt (2004).

Les principales caractéristiques des approches du rappel et des questions de base appliquées à l'EPA sont présentées au tableau 2. Pour plus de précisions, nous renvoyons le lecteur à Schouten (2007) et à Cobben et Schouten (2007). Dans l'approche du rappel, nous avons employé le questionnaire original visant les ménages administré par IPAO, tandis que dans l'approche des questions de base, nous avons utilisé un questionnaire abrégé dans des conditions de mode de collecte mixte. Le plan à mode mixte comportait une application en ligne, une application papier et une application ITAO. Cette dernière a été utilisée pour tous les ménages possédant un numéro de téléphone publié. Ceux n'ayant pas de numéro de téléphone publié ont reçu une lettre de présentation envoyée à l'avance, un questionnaire imprimé et la procédure d'entrée en communication avec un site Internet sécurisé contenant le questionnaire en ligne. Les répondants étaient libres de remplir le questionnaire imprimé ou le questionnaire en ligne.

Tableau 2
Caractéristiques des deux approches de l'étude de suivi

Approche du rappel	Approche des questions de base
• Questionnaire de l'EPA auquel devaient répondre tous les membres du ménage par IPAO. • 28 intervieweurs sélectionnés géographiquement parmi les intervieweurs qui, selon les antécédents, obtenaient les meilleurs résultats. • Intervieweur différent de celui qui avait reçu la non-réponse. • Formation supplémentaire concernant l'interaction au pas de la porte offerte aux intervieweurs. • Période de travail sur le terrain étendue de deux mois. • Intervieweur autorisé à offrir des mesures d'incitation. • Intervieweur susceptible de recevoir une prime. • Résumé imprimé des caractéristiques du ménage non répondant envoyé à l'intervieweur. • Affectation de l'adresse une semaine après la non-réponse.	• Questionnaire fortement condensé contenant les questions clés de l'EPA et auquel on peut répondre en une à trois minutes. • Plan de collecte des données à mode mixte (Internet, questionnaire papier et ITAO). • Questionnaire rempli par une personne par ménage, selon la méthode du prochain anniversaire. • Ménage contacté une semaine après qu'il ait été traité comme un cas de non-réponse.

La taille de l'échantillon de l'enquête pilote de l'EPA était de $n = 18\,074$ ménages, parmi lesquels 11 275 ont répondu. Les ménages non répondants ont été stratifiés en fonction de la cause de la non-réponse. Étaient admissibles au suivi les ménages qui n'avaient pas été traités, ceux avec

lesquels il n'y avait pas eu prise de contact et ceux qui avaient refusé de participer. Il a été jugé contraire à l'éthique de faire un suivi auprès des ménages qui n'avaient pas répondu pour d'autres raisons, telles qu'une maladie. En tout, 6 171 ménages étaient admissibles. Parmi ceux-ci, nous avons tiré deux échantillons aléatoires simples de taille 775. Dans les analyses, les ménages admissibles non échantillonnés ont été laissés de côté et ceux qui avaient été échantillonnés ont été pondérés en conséquence. Les 11 275 ménages ayant répondu à l'EPA et les 628 ménages inadmissibles ont tous reçu un poids de 1. Cela signifie que les probabilités d'inclusion sont inégales dans notre exemple.

Schouten (2007) a comparé les répondants à l'EPA aux non-répondants convertis et aux non-répondants persistants dans le cas de l'approche du rappel en utilisant un grand ensemble de caractéristiques démographiques et socioéconomiques. Pour prédire le type de réponse, il s'est servi de modèles de régression logistique. Il a conclu que les non-répondants convertis dans le cas de l'approche de rappel diffèrent des répondants à l'EPA en ce qui concerne les variables auxiliaires choisies. En outre, il n'a découvert aucune preuve que les non-répondants convertis différaient des non-répondants persistants en ce qui a trait aux mêmes caractéristiques. Ces constatations l'ont porté à conclure que la réponse combinée de l'EPA et de l'approche du rappel est plus représentative pour ce qui est des variables auxiliaires choisies.

Dans le cas de l'approche des questions de base, la réponse supplémentaire a été analysée par Cobben et Schouten (2007) en se servant du même ensemble de variables auxiliaires et des mêmes modèles de régression logistique. Dans le cas de cette forme de suivi, les constatations n'ont pas été les mêmes pour les ménages possédant et ne possédant pas de numéro de téléphone publié. Pour l'échantillon limité aux ménages ayant un numéro publié, ils ont obtenu les mêmes résultats que pour l'approche du rappel ; la réponse devient plus représentative après l'ajout des répondants ayant un numéro de téléphone publié qui ont répondu aux questions de base. Par contre, pour l'ensemble de la population, c'est-à-dire en incluant les ménages n'ayant pas de numéro de téléphone publié, l'inverse a été observé pour l'approche des questions de base. Celle-ci fournit « un plus grand nombre de résultats pareils » et, donc, accentue le contraste entre les répondants et les non-répondants. La combinaison des réponses à l'EPA et des réponses aux questions de base produit une composition moins représentative. Dans les modèles de régression logistique utilisés par Cobben et Schouten (2007), les indicateurs 0-1 pour la possession d'un numéro de téléphone publié et celle d'un travail rémunéré ont donné une contribution importante.

Cobben et Schouten (2007), ainsi que Schouten (2007) ont utilisé l'ensemble de variables auxiliaires énumérées au tableau 3. Ces variables ont été appariées à l'échantillon provenant de divers registres et données administratives. Elles ont été incluses dans les modèles de régression logistique pour les probabilités de réponse quand elles donnaient une contribution significative au seuil de signification de 5 %. Sinon, elles ont été exclues.

Tableau 3
Variables auxiliaires utilisées dans les études de Schouten (2007) et de Cobben et Schouten (2007). Le noyau du ménage est constitué du chef du ménage et de son (sa) partenaire si présent(e)

Variable
Le ménage possède un numéro de téléphone publié
Région du pays en 4 classes
Province et les 4 villes les plus grandes
Âge moyen en 6 classes
Groupe ethnique en 4 classes
Degré d'urbanisation en 5 classes
Type de ménage en 6 classes
Sexe
Valeur moyenne du logement au niveau du code postal en 11 classes
Au moins un membre du noyau du ménage est un travailleur autonome
Au moins un membre du noyau du ménage est inscrit au centre « Emploi et Revenu » (CWI)
Au moins un membre du noyau du ménage reçoit des allocations sociales
Au moins un membre du noyau du ménage possède un emploi rémunéré
Au moins un membre du noyau du ménage reçoit des allocations d'invalidité

Le tableau 4 donne la taille pondérée de l'échantillon, le taux de réponse, $\hat{R}$, $\mathrm{CI}_{0,05}^{\mathrm{BT}}$, $\hat{B}_m$ et $\hat{E}_m$ pour la réponse à l'EPA, la réponse à l'EPA combinée à la réponse au rappel, et la réponse à l'EPA combinée à la réponse aux questions de base. Les erreurs-types sont relativement grandes en ce qui concerne les études décrites aux sections suivantes, à cause de la pondération. La valeur de $\hat{R}$ augmente quand les répondants au rappel sont ajoutés aux répondants à l'EPA. Comme le taux de réponse et l'indicateur R augmentent tous deux, le biais absolu maximal $\hat{B}_m$ diminue. Les intervalles de confiance $\mathrm{CI}_{0,05}^{\mathrm{BT}}$ pour la réponse à l'EPA et pour les réponses combinées à l'EPA et au rappel se chevauchent. Cependant, l'application d'un test unilatéral mène au rejet de l'hypothèse nulle H_0: $R_{\mathrm{LFS}} - R_{\mathrm{LFS+CB}} \geq 0$ au niveau de signification de 5 %.

Tableau 4
Taille pondérée de l'échantillon, taux de réponse, indicateur R, intervalle de confiance, biais maximal et REQM maximale pour l'EPA, l'EPA plus le rappel, et l'EPA plus les questions de base pour l'ensemble étendu de variables auxiliaires

Réponse	n	Taux	$\hat{R}$	$\mathrm{CI}_{0,05}^{\mathrm{BT}}$	$\hat{B}_m$	$\hat{E}_m$
EPA	18 074	62,2 %	80,1 %	(77,5-82,7)	8,0 %	8,0 %
EPA + rappel	18 074	76,9 %	85,1 %	(82,4-87,8)	4,8 %	4,9 %
EPA + questions de base	18 074	75,6 %	78,0 %	(75,6-80,4)	7,3 %	7,3 %

Dans le tableau 4, nous notons une diminution de $\hat{R}$ quand nous comparons la réponse à l'EPA à la combinaison de cette réponse avec l'approche des questions de base. Cette diminution n'est pas significative. $\hat{B}_m$ diminue légèrement. Au tableau 5, cette comparaison est limitée aux ménages ayant un numéro de téléphone publié. En général, l'indicateur R est beaucoup plus élevé que pour l'ensemble des ménages. Comme la taille de l'échantillon est maintenant plus petite, les erreurs-types estimées sont plus grandes, comme en témoigne la largeur de l'intervalle de confiance. La valeur de $\hat{B}_m$ est plus faible. Pour la réponse combinée à l'EPA et à l'approche des questions de base, nous constatons un accroissement de $\hat{R}$, mais de nouveau, il n'est pas significatif. $\hat{B}_m$ diminue.

Tableau 5
Taille de l'échantillon, taux de réponse, indicateur R, intervalle de confiance, biais maximal et REQM maximale pour l'EPA et pour l'EPA plus les questions de base pour les ménages ayant un numéro de téléphone publié seulement et pour l'ensemble étendu de variables auxiliaires

Réponse	n	Taux	$\hat{R}$	$CI^{BT}_{0,05}$	$\hat{B}_m$	$\hat{E}_m$
EPA	10 135	68,5 %	86,3 %	(83,1-89,5)	5,0 %	5,1 %
EPA + questions de base	10 135	83,0 %	87,5 %	(84,3-90,7)	3,8 %	3,8 %

Dans le cas de l'exemple du suivi de l'EPA, nous constatons que les indicateurs R confirment les conclusions pour l'approche du rappel et celle des questions de base. En outre, l'accroissement de l'indicateur R consécutif à l'ajout des réponses au rappel est significatif au seuil de signification de 5 %.

5.3 Surveillance de la sécurité : essai pilote à mode mixte de 2006

En 2006, Statistique Pays-Bas a réalisé un essai pilote relatif à l'enquête de surveillance de la sécurité (*Safety Monitor*) en vue d'étudier les stratégies de collecte des données à mode mixte. Pour des renseignements détaillés, consulter Cobben, Janssen, Berkel et Brakel (2007). L'enquête de surveillance de la sécurité ordinaire, réalisée auprès de la population des Pays-Bas de 15 ans et plus, porte sur des questions relatives à la sécurité et à la façon dont la police remplit ses fonctions. Il s'agit d'une enquête à mode de collecte mixte. Les personnes dont le numéro de téléphone est publié sont approchées par ITAO. Celles qui ne peuvent pas être rejointes par téléphone sont approchées par IPAO. L'essai pilote de 2006 avait pour but d'évaluer la possibilité d'utiliser Internet comme l'un des modes dans une stratégie de collecte à mode mixte. Les personnes sélectionnées pour y participer ont d'abord été approchées au moyen d'une enquête en ligne. Les non-répondants à l'enquête en ligne ont été approchés de nouveau par ITAO s'ils possédaient un numéro de téléphone publié et par IPAO autrement. Le tableau 6 donne les taux de réponse pour l'enquête ordinaire, la réponse à l'essai pilote pour l'enquête en ligne uniquement et la réponse à l'essai pilote dans son ensemble. La réponse à l'enquête en ligne uniquement est faible. Seulement 30 % de personnes ont rempli le questionnaire en ligne. Cela signifie que près de 70 % des unités échantillonnées ont été réaffectées à l'IPAO ou à l'ITAO. Cela a produit une réponse supplémentaire d'environ 35 %. Le taux global de réponse est légèrement plus faible que celui observé pour l'enquête ordinaire.

Fouwels, Janssen et Wetzels (2006) ont procédé à une analyse univariée des compositions des réponses. Ils soutiennent que le taux de réponse est plus faible pour l'essai pilote, mais que cette diminution est relativement stable chez les divers sous-groupes démographiques. Ils constatent une baisse univariée du taux de réponse pour les variables auxiliaires d'âge, de groupe ethnique, de degré d'urbanisation et de type de ménage. Toutefois, ils relèvent des indices que la réponse devient moins représentative quand la comparaison est limitée aux personnes ayant répondu en ligne seulement. Cette constatation tient surtout pour l'âge des personnes échantillonnées, ce qui n'est pas étonnant.

Le tableau 6 donne la taille de l'échantillon, le taux de réponse, $\hat{R}$, $CI^{BT}_{0,05}$, $\hat{B}_m$ et $\hat{E}_m$ pour trois groupes : l'enquête ordinaire, l'enquête pilote limitée au mode de collecte en ligne et l'enquête pilote dans son ensemble. Les variables auxiliaires d'âge, de groupe ethnique, de degré d'urbanisation et de type de ménage ont été appariées d'après des registres et ont été sélectionnées dans le modèle logistique pour les probabilités de réponse. Le tableau 6 montre que l'indicateur R est plus faible pour la réponse en ligne que pour l'enquête ordinaire. La valeur p correspondante est proche de 5 %. Étant donné à la fois le faible taux de réponse et le faible indicateur R, le biais absolu maximal $\hat{B}_m$ est plus de deux fois plus grand que pour l'enquête ordinaire. Par contre, pour l'essai pilote dans son ensemble, l'indicateur R ainsi que $\hat{B}_m$ sont proches des valeurs observées pour l'enquête ordinaire. À cause de la plus petite taille de l'échantillon de l'essai pilote, les erreurs-types estimées sont plus grandes que pour l'enquête ordinaire.

Tableau 6
Taille de l'échantillon, taux de réponse, indicateur R, intervalle de confiance, biais maximal et REQM maximale pour la réponse à l'enquête de surveillance de la sécurité ordinaire, l'enquête pilote avec composantes en ligne uniquement et l'enquête pilote avec composantes en ligne et suivi par IPAO/ITAO

Réponse	n	Taux	$\hat{R}$	$CI^{BT}_{0,05}$	$\hat{B}_m$	$\hat{E}_m$
Ordinaire	30 139	68,9 %	81,4 %	(80,3-82,4)	6,8 %	6,8 %
Pilote – en ligne	3 615	30,2 %	77,8 %	(75,1-80,5)	18,3 %	18,4 %
Pilote – en ligne plus	3 615	64,7 %	81,2 %	(78,3-84,0)	7,3 %	7,4 %

Les résultats présentés au tableau 6 ne contredisent pas ceux de Fouwels et coll. (2006). Nous constatons aussi que la composition de la réponse en ligne à l'essai pilote est moins équilibrée, tandis que celle de la réponse complète à l'essai pilote n'est pas sensiblement moins bonne que celle de la réponse à l'enquête de surveillance de la sécurité proprement dite.

6. Discussion

Nous poursuivons trois grands objectifs dans le présent article : donner une définition mathématiquement rigoureuse et une perception de la représentativité de la réponse, construire un indicateur possible de la représentativité et illustrer empiriquement l'utilisation de l'indicateur. Comme nous l'avons vu, l'indicateur proposé est un exemple de ce que nous appelons les indicateurs R, où « R » signifie représentativité. Au moyen de l'exemple empirique, nous cherchons à appuyer l'idée que ces indicateurs R sont des outils valables pour comparer différentes enquêtes ou stratégies de collecte des données. Les indicateurs R sont utiles s'ils permettent de confirmer les résultats d'analyses complexes effectuées dans le cadre d'études ayant trait à des enquêtes réalisées sur de multiples périodes ou à de multiples enquêtes portant sur un sujet particulier.

L'indicateur R décrit dans le présent article est prometteur, parce qu'il est facile à calculer et peut être interprété et normalisé quand les propensions à répondre peuvent être estimées sans erreur. L'application à des données d'enquête réelles montre que l'indicateur R confirme les analyses antérieures de la composition de la non-réponse. Naturellement, d'autres indicateurs R peuvent être construits en choisissant simplement différentes fonctions de distance entre les vecteurs des propensions à répondre. L'indicateur R et les graphiques présentés dans l'article peuvent être calculés en utilisant la plupart des progiciels statistiques standard.

Le calcul des indicateurs R est fondé sur un échantillon et s'appuie sur des modèles pour les propensions à répondre individuelles. Donc, ces indicateurs sont eux-mêmes des variables aléatoires et leur estimation comporte deux étapes qui influencent leur biais et leur variance. Cependant, c'est surtout la modélisation des propensions à répondre qui a des conséquences importantes. Le fait de se limiter à l'échantillon pour estimer les indicateurs R implique que ceux-ci sont moins précis, mais asymptotiquement, cette restriction n'introduit pas de biais. Le choix et l'ajustement du modèle sont habituellement effectués en imposant un niveau de signification et en ajoutant uniquement dans le modèle les interactions dont la contribution est significative. Ce dernier point signifie que la taille de l'échantillon et la disponibilité de données sur les variables auxiliaires jouent un rôle important dans l'estimation des propensions à répondre. La stratégie de sélection du modèle peut introduire un biais. Diverses approches évidentes peuvent être utilisées pour traiter la dépendance à l'égard de la taille de l'échantillon. On pourrait ne pas sélectionner un modèle, mais fixer une stratification d'avance. De cette façon, on évite d'introduire un biais, mais les erreurs-types ne sont pas contrôlées et peuvent être considérables. On pourrait aussi prendre la validation empirique comme point de départ pour élaborer les « pratiques exemplaires » pour les indicateurs R.

Nous avons appliqué l'indicateur R proposé à deux études réalisées à Statistique Pays-Bas ces dernières années et dont les résultats avaient été examinés en détail par d'autres auteurs. L'accroissement ou la diminution de la valeur de l'indicateur R concorde avec les résultats des analyses plus détaillées effectuées par ces auteurs. Par conséquent, nous concluons que les indicateurs R peuvent être des outils valables. Cependant, un plus grand nombre de données empiriques sont manifestement nécessaires.

L'application de l'indicateur R a révélé qu'il n'existe aucune relation claire entre le taux de réponse et la représentativité de la réponse. Les taux de réponse plus élevés ne mènent pas nécessairement à une réponse plus équilibrée. Naturellement, nous constatons que les taux de réponse plus élevés réduisent le risque d'un biais de non-réponse. Plus le taux de réponse est élevé, plus le biais absolu maximal observé pour les items d'intérêt est faible.

L'application aux études choisies montre que les erreurs-types diminuent quand la taille de l'échantillon augmente, comme il fallait s'y attendre, mais elles demeurent relativement grandes pour des tailles d'échantillon modestes. Par exemple, pour une taille d'échantillon de 3 600, nous observons une erreur-type d'environ 1,3 %. Donc, si nous émettons l'hypothèse d'une loi normale, l'intervalle de confiance à 95 % a une largeur approximative de 5,4 %. La taille de l'échantillon de l'EPA est d'environ 30 000 unités. L'erreur-type est de l'ordre de 0,5 % et l'intervalle de confiance à 95 % correspondant à une largeur d'environ 2 %. Les erreurs-types sont plus grandes que nous ne l'avions pensé.

Le présent article contient une première étude empirique d'un indicateur R et de son erreur-type. Des travaux de recherche théoriques et empiriques beaucoup plus approfondis seront nécessaires pour comprendre pleinement les indicateurs R et leurs propriétés. En premier lieu, nous n'avons pas du tout tenu compte des items de l'enquête. Manifestement, il est absolument nécessaire que nous le fassions dans l'avenir. Cependant, comme nous l'avons déjà soutenu, les indicateurs R dépendent de l'ensemble de variables auxiliaires. Nous pouvons, par conséquent, conjecturer que, comme dans le cas des méthodes de correction de la non-réponse, la mesure dans laquelle les indicateurs R

prédisent le biais de non-réponse des variables étudiées dépend du mécanisme de création de données manquantes. Si celui-ci est fortement non ignorable, les indicateurs R ne donneront pas de bons résultats. Cependant, sans aucun renseignement sur le mécanisme de données manquantes, aucun autre indicateur ne le ferait non plus. C'est pourquoi nous avons établi la notion de biais absolu maximal, car elle donne une limite au biais de non-réponse sous le pire scénario. Un deuxième sujet de futurs travaux de recherche est le calcul théorique de l'erreur-type de l'indicateur R utilisé dans le présent article. Les erreurs obtenues par le bootstrap non paramétrique ne donnent que des approximations naïves. Or, si nous voulons que les indicateurs R jouent un rôle plus actif dans la comparaison de diverses stratégies, nous devons obtenir des formes (approximativement) explicites. Troisièmement, nous devrions étudier la relation entre le choix et le nombre de variables auxiliaires, d'une part, et les erreurs-types de l'indicateur R, d'autre part.

Remerciements

Les auteurs remercient Bob Groves, Björn Janssen, Geert Loosveldt, ainsi que le rédacteur associé et deux examinateurs de leurs suggestions et commentaires constructifs.

Bibliographie

Agresti, A. (2002). Categorical data analysis. *Wiley Series in Probability and Statistics*. New York : John Wiley & Sons, Inc., NY, USA.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 2, 238-246.

Bertino, S. (2006). A measure of representativeness of a sample for inferential purposes. *Revue Internationale de Statistique*, 74, 149-159.

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, 3, 251-260.

Cobben, F., Janssen, B., Berkel, K. van et Brakel, J. van den (2007). Statistical inference in a mixed-mode data collection setting. Document présenté au ISI 2007, 23 au 29 août 2007, Lisbon, Portugal.

Cobben, F., et Schouten, B. (2007). An empirical validation of R-indicators. Papier de discussion, CBS, Voorburg.

Curtin, R., Presser, S. et Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, 64, 413-428.

Efron, B., et Tibshirani, R.J. (1993). An introduction to the bootstrap. Chapman & Hall/CRC.

Fouwels, S., Janssen, B. et Wetzels, W. (2006). Experiment mixed-mode waarneming bij de VMR. Rapport technique SOO-2007-H53, CBS, Heerlen.

Goodman, L.A., et Kruskal, W.H. (1979). Measures of association for cross-classifications. Springer-Verlag, Berlijn, Duitsland.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70, 5, 646-675.

Groves, R.M., et Peytcheva, E. (2006). The impact of nonresponse rates on nonresponse bias: A meta-analysis. Document présenté au 17th International Workshop on Household Survey Nonresponse, 28 au 30 août, Omaha, NE, USA.

Groves, R.M., Presser, S. et Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68, 2-31.

Hájek, J. (1981). Sampling from finite populations. New York : Marcel Dekker, USA.

Hansen, M.H., et Hurwitz, W.H. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.

Heerwegh, D., Abts, K. et Loosveldt, G. (2007). Minimizing survey refusal and noncontact rates: Do our efforts pay off? *Survey Research Methods*, 1, 1, 3-10.

Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 2, 119-127.

Keeter, S., Miller, C., Kohut, A., Groves, R.M. et Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, 64, 125-148.

Kersten, H.M.P., et Bethlehem, J.G. (1984). Exploring an reducing the nonresponse bias by asking the basic question. *Statistical Journal of the United Nations*, ECE 2, 369-380.

Kohler, U. (2007). Surveys from inside: An assessment of unit nonresponse bias with internal criteria. *Survey Research Methods*, 1, 2, 55-67.

Kruskal, W., et Mosteller, F. (1979a). Representative sampling I: Non-scientific literature. *Revue Internationale de Statistique*, 47, 13-24.

Kruskal, W., et Mosteller, F. (1979b). Representative sampling II: Scientific literature excluding statistics. *Revue Internationale de Statistique*, 47, 111-123.

Kruskal, W., et Mosteller, F. (1979c). Representative sampling III: Current statistical literature. *Revue Internationale de Statistique*, 47, 245-265.

Little, R.J.A., et Rubin, D.B. (2002). Statistical analysis with missing data. Wiley Series in Probability and Statistics. New York : John Wiley & Sons, Inc., NY, USA.

Marsh, H.W., Balla, J.R. et McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 3, 391-410.

Merkle, D.M., et Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. Dans *Survey Nonresponse* (Éds. R.M. Groves, D.A. Dillman, J.L. Eltinge et R.J.A Little). New York : John Wiley & Sons, Inc., 243-258.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

Särndal, C., et Lundström, S. (2005). Estimation in Surveys with Nonresponse. Wiley Series in Survey Methodology, John Wiley & Sons, Chichester, England.

Särndal, C., Swensson, B. et Wretman, J. (2003). Model-assisted survey sampling. Springer Series in Statistics, Springer, New York.

Schouten, B. (2004). Adjustment for bias in the Integrated Survey on Household Living Conditions (POLS) 1998. Papier de discussion 04001, CBS, Voorburg, disponible au site Web http://www.cbs.nl/nl-NL/menu/methoden/research/discussionpapers/archief/2004/default.htm.

Schouten, B., et Cobben, F. (2007). R-indicators for the comparison of different fieldwork strategies and data collection modes, Article de discussion 07002, CBS, Voorburg. Disponible au http://www.cbs.nl/nl-NL/menu/methoden/research/discussionpapers/archief/2007/default.htm.

Stoop, I. (2005). Surveying nonrespondents. *Field Methods*, 16, 23-54.

Voogt, R. (2004). I am not interested: Nonresponse bias, response bias and stimulus effects in election research. Thèse de doctorat, University of Amsterdam, Amsterdam.