



N° 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

**Symposium 2004 : Méthodes
innovatrices pour enquêter
auprès des populations
difficiles à joindre**

2004



ANALYSE À VARIABLES MULTIPLES DE LA NON-RÉPONSE DANS LES MINORITÉS ETHNIQUES

Remco Feskens, Joop Hox, Gerty Lensvelt-Mulders et Hans Schmeets¹

RÉSUMÉ

Dans cet article, nous examinons la non-réponse dans les minorités ethniques aux Pays-Bas. Dans le logiciel Mplus, nous avons élaboré un modèle par équation structurelle comportant une diversité de résultats de réponse. Nous avons pris en compte l'incidence de l'ethnicité sur ces résultats pour diverses variables socioéconomiques et sociodémographiques. Nous avons constaté que l'effet de réponse était presque entièrement expliqué par le degré d'urbanisation. Par des analyses en groupes multiples, nous avons aussi examiné les différences entre groupes ethniques pour les prédicteurs de résultats de réponse. Là encore, nous avons constaté que l'urbanisation influe négativement sur les probabilités de réponse de tous les groupes ethniques et, plus particulièrement, sur les probabilités de prise de contact. Il reste que cet effet négatif est un peu plus marqué dans le cas des groupes d'origine étrangère non occidentale.

MOTS CLÉS : Minorités ethniques, modélisation par équation structurelle, non-réponse.

1. INTRODUCTION

Ces dernières années, les taux de non-réponse se sont élevés dans presque tous les pays d'Occident (De Heer et De Leeuw, 2002). Cette non-réponse pose un problème pour plusieurs raisons. Premièrement, elle a un effet réducteur sur le nombre de répondants et donc sur la précision des estimations. Deuxièmement, si la non-réponse est sélective, les estimations d'enquête peuvent s'en trouver biaisées et ne rendront pas fidèlement compte des « valeurs vraies » de la population visée (Thornberry et Massey, 1988). Il y a non-réponse sélective lorsque les non-répondants diffèrent systématiquement des répondants d'une manière significative pour les objectifs de l'enquête en question (Groves et Couper, 1998). Troisièmement, la non-réponse peut accroître les coûts de la recherche par enquête, puisqu'il faut redoubler d'efforts pour parvenir à la taille d'échantillon recherchée (Hox et De Leeuw, 1998). La conséquence la plus fâcheuse de la non-réponse est le biais introduit dans les estimateurs ponctuels (Groves, 1989). Elle est plus probable lorsque, dans la société, des groupes particuliers présentent des taux de réponse inférieurs à la moyenne, d'où une plus grande probabilité que les non-répondants diffèrent systématiquement des répondants, la non-réponse perdant tout caractère aléatoire. Pour y aller de conjectures au sujet du biais de non-réponse, il est bon de regarder les taux de réponse parmi les sous-populations (Thornberry et Massey, 1988).

Pour diverses raisons, notamment des taux de non-réponse supérieurs à la moyenne, Statistique Pays-Bas a de la difficulté à faire enquête auprès de cinq sous-populations qualifiées de populations difficiles à joindre (Snijkers, 2003). Une de ces populations est la population immigrante (minorités ethniques) qui forme environ 20 % de la société néerlandaise (Statistique Pays-Bas). Dans notre pays, nous définissons le membre d'une minorité ethnique comme quiconque habite aux Pays-Bas et a un père et/ou une mère nés à l'étranger (Reep, 2003). Nous distinguons habituellement en plus les gens originaires d'Occident (le père et/ou la mère étant nés dans un autre pays d'Europe, en Amérique du Nord, en Australie, au Japon ou en Indonésie) des gens dont l'origine est extraoccidentale (il s'agit principalement des Turcs, des Marocains, des Surinamiens et des Antillais)². Ne serait-ce que pour cette large définition que nous employons, nous constatons que presque un habitant des Pays-Bas sur cinq appartient à une minorité ethnique. Environ la moitié sont des étrangers non occidentaux (c'est là 10 % de la population totale). Presque 9 % de la population néerlandaise est d'extraction étrangère occidentale. Les problèmes de réponse ne sont

¹ R.C.W. Feskens, J. Hox, G. Lensvelt-Mulders (Université d'Utrecht) et J.J.G. Schmeets (Statistique Pays-Bas), Heidelberglaan 1, 3584 CS Utrecht, Pays-Bas (r.c.w.feskens@fss.uu.nl). Les vues exposées dans le présent document appartiennent aux auteurs et ne correspondent pas nécessairement aux politiques de Statistique Pays-Bas.

² Par souci de simplicité, nous parlons d'une origine étrangère occidentale et extraoccidentale dans le présent article.

pas propres aux Pays-Bas en ce qui concerne les minorités ethniques, celles-ci présentant des taux de réponse réduits dans presque tous les pays d'Occident (Feskens, 2005).

Notons en particulier que la population d'origine étrangère extraoccidentale a généralement des taux de réponse inférieurs à ceux de la population de souche, mais que, par ailleurs, elle vit surtout en région urbaine, connaît plus souvent le chômage et est moins scolarisée que la population de souche. Ces caractéristiques sont en corrélation négative avec la réponse (Groves et Couper, 1998; Stoop, 2004). Dans une étude antérieure par modélisation de régression loglinéaire, on a démontré que les hauts taux de non-réponse des minorités ethniques étaient imputables à la situation socioéconomique et à l'urbanisation (Schmeets et Michiels, 2003). Dans la présente étude, nous voulons voir si cette conclusion s'impose tout autant lorsque nous employons les techniques de modélisation par équation structurelle. Il s'agira en outre pour nous d'élaborer ce modèle en y incluant plus de variables ainsi que des effets d'interaction. Nous indiquerons enfin les effets produits sur l'absence de prise de contact et le refus d'interview.

Ces considérations nous mènent à trois questions de recherche :

- 1) À quel point la non-réponse pose-t-elle un problème si on considère l'absence de réponse, l'absence de prise de contact et le refus dans le cas des minorités ethniques néerlandaises?
- 2) Quelle est l'incidence de l'ethnicité sur les différences de résultats de réponse si on prend en compte d'autres variables socioéconomiques et sociodémographiques?
- 3) Quelles variables ou prédispositions sont-elles liées au phénomène de l'absence de réponse ou de prise de contact et du refus dans les groupes ethniques?

2. DONNÉES

Nos analyses ont porté sur les fichiers de l'enquête permanente sur les conditions de vie (« Permanent Survey on Living Conditions » ou POLS). C'est une enquête réalisée par Statistique Pays-Bas. En général, 40 000 interviews sont effectuées par an. Les unités d'observation sont les individus. La base de sondage exploitée est le registre démographique de toutes les administrations municipales de base au pays. La POLS comporte un échantillonnage en grappes. On échantillonne d'abord les collectivités et ensuite les gens. Les grandes villes sont automatiquement incluses (Schouten, 2003). La participation est volontaire. Chaque mois, on prélève un échantillon d'environ 3 500 personnes. En 1998, il y avait deux mois de travail de terrain pour les 12 échantillons consécutifs. Le premier mois, Statistique Pays-Bas a procédé à une collecte IPAO. Dans ce cas, on a relancé par technique ITAO les non-répondants dont on savait qu'ils avaient le téléphone. Les non-répondants sans ligne téléphonique connue et les unités d'échantillonnage incapables de collaborer à cause de la maladie ont été relancés en mode IPAO (Schouten, 2003).

On a complété les données d'enquête POLS par des données du registre démographique et des renseignements sur l'emploi et les prestations sociales (Schmeets et Michiels, 2003; Schouten, 2003). Grâce à ce jumelage de dossiers administratifs, nous disposons de données sociodémographiques et socioéconomiques sur les non-répondants au niveau des codes postaux. Dans l'enquête POLS de 1998, on a échantillonné 39 431 unités. Selon la définition de la réponse de l'AAPOR, le nombre de répondants s'est établi à 23 993 ou 61,4 % (AAPOR, 2000), ce qui n'est pas spécialement bas aux Pays-Bas (voir, par exemple, De Heer, 1999).

L'information complémentaire obtenue a permis d'étudier les non-répondants. On ne disposait toutefois pas de ce complément d'information pour toutes les unités échantillonnées. Les données manquaient pour 1 143 des unités (2,9 % de tout l'échantillon). Supposant que ces unités manquaient au hasard, on les a retranchées du fichier de données. Comme il s'agit de valeurs plutôt modestes, il est possible de retirer ces unités de l'échantillon sans grande perte d'information. Comme nous avons uniquement analysé les unités échantillonnées de 15 à 65 ans, il nous est resté 28 542 unités pour l'analyse. Dans ce sous-échantillon, la réponse est un peu moins importante, à savoir 60,4 %, ce qui correspond à 17 123 unités. Pour ne pas s'en remettre à la chance, on a divisé aléatoirement ce nouveau fichier en un fichier d'exploration et un fichier de validation. Le premier était constitué des cas impairs (14 101 cas) et le second, des cas pairs (14 271).

3. MÉTHODES

Nous nous sommes d'abord attardés aux relations bidimensionnelles entre les groupes ethniques et plusieurs catégories de réponse. Grâce à cette information, nous avons pu répondre à la première question de recherche. Pour répondre à la deuxième, nous avons effectué une modélisation par équation structurelle dans le progiciel Mplus, ce qui nous a permis de combiner des variables latentes et des relations structurelles entre celles-ci et d'autres variables observées (Kline, 1998). Avec Mplus, nous avons pu élaborer un modèle par équation structurelle avec une variable dépendante dichotomique. Comme l'indiquent Groves et Couper (1998), la distinction des cas d'absence de prise de contact, de refus et autres dans le phénomène de la non-réponse nous amène à considérer les causes diverses de chaque résultat. Ainsi, nous avons analysé non seulement les résultats de réponse, mais aussi les cas de non-prise de contact et de refus, d'où un enrichissement de notre information. Les autres causes de non-réponse n'avaient que peu d'influence sur le taux de réponse, comme on peut le voir au tableau 1.

Nous avons également voulu savoir pour quelles variables les deux groupes (population de souche et d'origine étrangère occidentale et population d'origine extraoccidentale) différaient sur le plan de la réponse. Ce faisant, nous avons évalué l'effet d'interaction de l'ethnicité par une analyse en groupes multiples où nous comparions les « coefficients de direction » de ces groupes. Là encore, nous avons distingué les résultats de réponse et avons pu examiner quelles prédispositions de réponse existent parmi les sous-groupes. Comme l'échantillon était de grande taille, nous avons évalué l'ajustement du modèle par deux indices moins sensibles à la taille d'échantillon, à savoir l'indice comparatif d'ajustement de Bentler (ICA) et l'erreur quadratique moyenne de l'approximation (EQMA).

4. RÉSULTATS

4.1 À quel point la non-réponse (absence de réponse ou de prise de contact et refus) pose-t-elle un problème dans les minorités ethniques des Pays-Bas?

Le taux général de réponse s'est établi à 60,4 % pour les unités échantillonnées de 15 à 65 ans dans l'enquête POLS de 1998. La décomposition des résultats de réponse donne le tableau suivant pour les trois groupes ethniques considérés :

Tableau 1 : Résultats en pourcentage parmi les groupes ethniques visés par l'enquête POLS de 1998

	Population de souche	Population d'origine étrangère occidentale	Population d'origine étrangère extraoccidentale
Réponse	62,0	57,2	39,7
Absence de prise de contact	12,0	15,9	26,4
Refus	24,7	23,0	20,1
Problèmes de langue	0,0	2,5	13,0
Autres	1,3	1,8	1,0

L'addition des chiffres peut ne pas donner 100 % à cause des arrondis.

La distribution des résultats de non-réponse est plus ou moins la même pour la population d'extraction étrangère occidentale que pour la population de souche. Le nombre relatif de cas d'absence de prise de contact et de non-réponse à cause de l'obstacle de la langue est plus élevé dans la population d'origine étrangère extraoccidentale que dans la population de souche, résultat déjà obtenu dans d'autres enquêtes (Feskens, 2005). C'est l'explication des taux de réponse inférieurs de la population d'origine étrangère non occidentale.

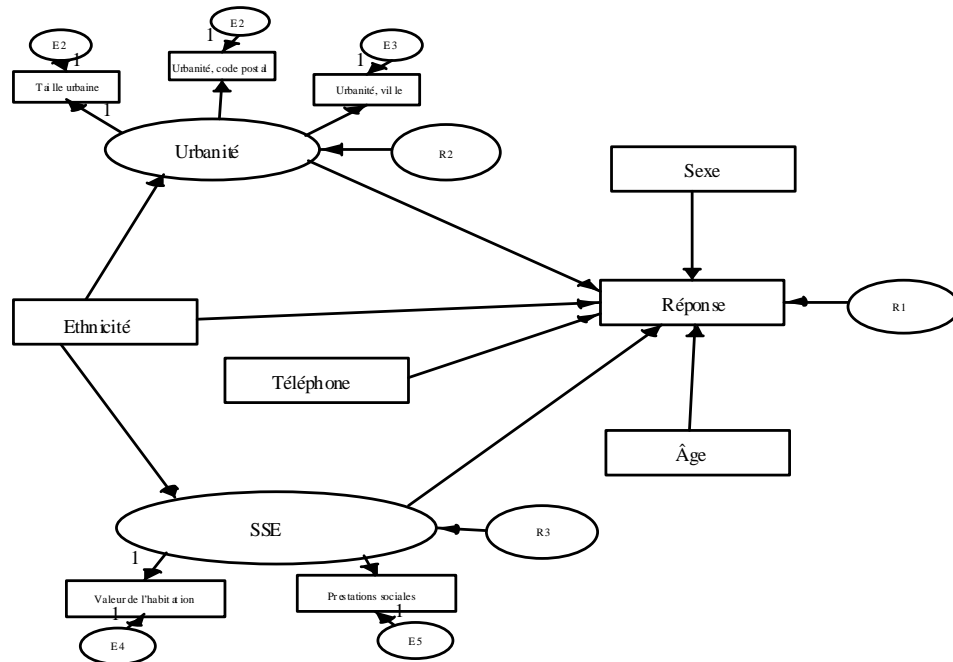
Nous avons regardé les relations bidimensionnelles entre l'ethnicité et diverses variables sociodémographiques et socioéconomiques. Là encore, les caractéristiques relevées de la population d'origine étrangère occidentale ressemblaient fort à celles de la population de souche, mais les différences étaient marquées dans le cas de la population d'origine extraoccidentale. C'est pourquoi nous avons voulu concentrer notre analyse sur cette dernière population.

4.2 Quelle est l'incidence de l'ethnicité sur les résultats de réponse après prise en compte d'autres variables socioéconomiques et sociodémographiques?

4.2.1 Modélisation par équation structurelle

Nous désirions savoir si la non-réponse subissait toujours l'influence de l'ethnicité une fois prise en compte les variables sociodémographiques et socioéconomiques. Nous avons d'abord regardé les relations bidimensionnelles entre chacune de ces variables et la réponse de tout groupe ethnique pris isolément. Nous avons examiné ces mêmes relations entre l'ensemble des groupes ethniques et ces mêmes variables. Munis d'une telle information et d'un cadre théorique, nous avons élaboré un modèle par équation structurelle où il serait possible de tenir compte de l'incidence de l'ethnicité sur différents résultats de réponse pour d'autres variables (voir la figure 1).

Figure 1 : Modélisation par équation structurelle



Le volet de mesure du modèle visait l'urbanité et la situation socioéconomique en variables latentes. L'urbanité en variable latente était mesurée par l'urbanisation en variable observée au niveau des codes postaux, le degré d'urbanisation de la ville et la taille urbaine. Ces variables sont suffisamment en corrélation les unes avec les autres, mais on ne peut parler de corrélation dans la mesure où elles appréhendent le même phénomène. La SSE en variable latente a été mesurée par la valeur de l'habitation et des indicateurs de réception de prestations sociales. À des fins de caractérisation, nous avons fixé à l'unité les « charges factorielles » des indicateurs de l'urbanité de la ville et de la valeur de l'habitation.

Nous avons estimé des régressions en probits pour les indicateurs catégoriques et des régressions linéaires simples pour les indicateurs continus (Muthén, 1998-2004). Si nous avons élaboré les deux variables latentes, c'est que l'exploitation de l'information supplémentaire de tous les indicateurs rendait possible la caractérisation de l'urbanité et de la situation socioéconomique avec plus de précision que prévu. Ajoutons que certains indicateurs font l'objet d'une caractérisation subjective. En disposant de variables latentes avec plus d'indicateurs, on se trouve à réduire cette forme d'erreur de mesure. Les relations entre les autres variables observées et les variables latentes pour la réponse ont constitué le volet structurel de notre modèle. Les variables d'observation sont l'ethnicité, l'existence connue du téléphone par réseau terrestre, le sexe et l'âge. Nous avons opéré la régression de ces variables observées sur la variable binaire de réponse.

Nous désirions savoir si la relation entre l'ethnicité et la réponse était influencée par la SSE et l'urbanité. Nous avons donc mis en régression l'ethnicité par rapport à ces variables latentes. Nous supposons que les relations étaient unidirectionnelles et que les variables latentes de la situation socioéconomique et de l'urbanité et les variables de la SSE et de la desserte téléphonique seraient en covariation. Nous avons estimé les coefficients de régression par les moindres carrés sans pondération, ce qui peut s'interpréter comme des coefficients en probits. Les estimations figurant entre parenthèses sont des coefficients normalisés où interviennent les variances tant des variables latentes continues que des variables de contexte et de résultat (Muthén et Muthén, 1998-2004). Le signe des coefficients de régression indique pour quel degré de cette caractéristique les probabilités de non-réponse (signe négatif) ou de réponse (sans signe) s'infléchissent. Nous avons posé que les erreurs de mesure n'étaient pas en corrélation.

4.2.2 Réponse

Le tableau 2 livre les résultats de la modélisation par équation structurelle pour la variable dépendante réponse.

Tableau 2 : Modélisation par équation structurelle avec comme variable dépendante réponse

	Variables expliquées					
	Estimations relatives à la réponse	Erreurs-types	Estimations relatives à l'urbanité	Erreurs-types	Estimations relatives à la SSE	Erreurs-types
Prédicteurs						
Taille urbaine			1,000 (0,874)fixe	0,000		
Urbanité, ville			0,825 (0,914)**	0,011		
Urbanité, code postal			0,870 (0,887)**	0,015		
Valeur de l'habitation					1,000 (0,648)fixe	0,000
Prestations sociales					-0,370 (-0,491)**	0,020
Sexe	0,075 (0,037)**	0,021				
Âge	-0,002 (-0,033)**	0,001				
Urbanité	-0,097 (-0,142)**	0,011				
SSE	0,047 (0,063)**	0,018				
Téléphone	0,341 (0,139)**	0,031				
Ethnicité	-0,285 (-0,072)**	0,048	1,453 (0,250)**	0,054	-1,805 (-0,340)**	0,073

Note : Variable dépendante avec le code 1 pour la réponse et le code 0 pour la non-réponse.

**p<0,01

Chi carré : 125,432 (dl= 4); EQMA : 0,046; ICA : 0,972; R² : 0,072.

Fichier de validation : Chi carré : 91,611 (dl= 3); EQMA : 0,045; ICA : 0,980; R² : 0,082.

La relation entre l'ethnicité et la réponse est presque entièrement expliquée par l'urbanité et la situation socioéconomique des unités échantillonnées. L'incidence négative de l'ethnicité sur la réponse est largement expliquée par l'urbanité. La régression normalisée en probits de l'urbanité sur la réponse présente une valeur de -0,142 et l'effet normalisé de l'ethnicité sur l'urbanité s'établit à 0,250. L'urbanité et la desserte téléphonique ont une grande incidence en particulier sur les probabilités de réponse. La SSE en variable latente n'influe guère sur ces probabilités.

Les minorités ethniques ont des taux de réponse inférieurs à ceux de la population de souche, mais elles sont aussi d'une présence disproportionnée dans les agglomérations urbaines. Ces résultats font voir que l'effet d'urbanité « cause » plus particulièrement cette faiblesse relative des taux de réponse dans les minorités ethniques, et non pas l'ethnicité proprement dite. Après prise en compte des autres variables, l'ethnicité a seulement une incidence légère sur les probabilités de réponse et un coefficient normalisé plutôt modeste de régression de l'ethnicité sur la réponse (-0,072) reste dans ce cadre d'analyse à plusieurs variables.

La situation socioéconomique des unités échantillonnées n'influe pas outre mesure sur les probabilités de réponse, d'où l'impression que l'effet négatif de l'ethnicité sur la réponse est principalement influencé par les variables sociodémographiques plutôt que par les caractéristiques socioéconomiques. Cette constatation vaut aussi pour la population de souche, puisque, comparativement à la population non urbaine, la population urbaine présente des probabilités inférieures de réponse sans égard à l'ethnicité ni à la situation socioéconomique. L'existence du téléphone par réseau terrestre s'avère un indicateur révélateur de la réponse, et ce, non seulement pour la seconde partie des opérations de terrain où les stratégies de relance diffèrent entre le mode ITAO pour les non-répondants

dont on sait qu'ils ont le téléphone et le mode IPAO dans les autres cas, mais aussi pour la première étape où il n'y a pas d'appels téléphoniques. L'âge et le sexe sont seulement d'une faible incidence sur la réponse. Pour les unités échantillonnées plus âgées, les probabilités de réponse sont un peu moindres.

L'incidence de l'ethnicité décroît nettement après prise en compte des autres variables. Non seulement, le « coefficient de direction » évolue en baisse après prise en compte de l'indicateur de l'ethnicité, mais le pseudo-R² augmente seulement à proportion de 1,0 % (de 6,2 % à 7,2 %). De faibles coefficients de direction et pseudo-R² semblent indiquer que prédire la réponse est toujours assez difficile, même si on dispose d'une riche information de contexte. On est aussi porté à penser que le biais de réponse n'est pas aussi marqué que certaines relations bidimensionnelles ne semblent le montrer. Il se peut toutefois qu'une certaine non-réponse sélective se manifeste dans les régions urbaines. La population urbaine est quelque peu sous-représentée dans cette enquête, quelle qu'en soit l'ethnicité.

4.2.3 Prise de contact

Nous présentons au tableau 3 les résultats de la modélisation par équation structurelle pour la prise de contact comme variable dépendante.

Tableau 3 : Modélisation par équation structurelle avec la variable dépendante de la prise de contact

	Variables expliquées					
	Estimations relatives à la prise de contact	Erreurs-types	Estimations relatives à l'urbanité	Erreurs-types	Estimations relatives à la SSE	Erreurs-types
Prédicteurs						
Taille urbaine			1,000 (0,874)fixe	0,000		
Urbanité, ville			0,825 (0,914)**	0,011		
Urbanité, code postal			0,870 (0,887)**	0,015		
Valeur de l'habitation					1,000 (0,649)fixe	0,000
Prestations sociales					-0,368 (-0,489)**	0,020
Sexe	0,060 (0,030)**	0,027				
Âge	0,004 (0,056)**	0,001				
Urbanité	-0,183 (-0,267)**	0,012				
SSE	-0,014 (-0,019)ns	0,020				
Téléphone	0,386 (0,156)**	0,035				
Ethnicité	-0,215 (-0,054)**	0,052	1,453 (0,250)**	0,054	-1,806 (-0,340)**	0,073

Note : Variable dépendante avec le code 1 pour la prise de contact et le code 0 pour l'absence de prise de contact.

**p<0,01; ns : non significatif

Chi carré : 123,290 (dl= 4); EQMA : 0,046; ICA : 0,972; R² = 0,104.

Fichier de validation : Chi carré : 90,715 (dl= 3); EQMA : 0,045; ICA : 0,980; R² : 0,113.

Le rôle de l'urbanité est encore plus prononcé en ce qui concerne la prise de contact. Le coefficient normalisé de régression de l'urbanité sur les probabilités de prise de contact est de -0,267. L'effet de l'ethnicité sur la prise de contact diminue par rapport à son effet sur le taux de réponse, indice que cet effet est encore plus influencé par l'urbanité. Entre autres auteurs, Groves et Couper (1998) ont fait remarquer qu'il est plus difficile de prendre contact avec les unités échantillonnées en région urbaine qu'en milieu non urbain.

D'après les résultats considérés, cette constatation vaut aussi pour les minorités ethniques. Dans ces groupes minoritaires, la non-réponse est largement déterminée par la faiblesse des taux de prise de contact. Cette faiblesse n'est pas propre aux minorités ethniques, étant largement en médiation par l'urbanité. On voit aussi pourquoi ces minorités présentent des taux de réponse inférieurs. Les difficultés de prise de contact sont avant tout concentrées en région urbaine. On a relancé dans le second mois en mode ITAO les non-répondants dont on savait qu'ils avaient le téléphone. On a relancé les autres répondants en reprenant le mode IPAO. Il reste que, en raison de l'insuffisance des capacités d'interview, on n'a pas relancé tous les non-répondants n'ayant pas le téléphone ou il y a eu moins de tentatives de prise de contact dans ce cas que dans celui des non-répondants en relance ITAO. On ne s'étonnera donc pas que le coefficient de régression de l'existence du téléphone sur les probabilités de prise de contact soit élevé. On a néanmoins relevé cet effet positif de l'existence connue du téléphone sur les probabilités de prise de contact dans le premier mois aussi, où il n'y a pas eu d'appels téléphoniques. Enfin, les résultats indiquent que les

femmes et les personnes âgées sont un peu plus faciles à joindre, bien que les probabilités de prise de contact demeurent faibles dans leur cas.

4.2.4 Refus

Les non-répondants qui ne sont pas joints n'ont pas la possibilité de refuser de participer à l'enquête (Hox et De Leeuw, 1998). Dans notre analyse des unités échantillonnées qui ont refusé de participer, nous nous sommes donc limités aux unités échantillonnées jointes au départ. Il nous est resté 12 202 unités échantillonnées pour le fichier d'exploration; le fichier de validation comptait 12 366 de ces unités. On trouvera les résultats au tableau 4.

Tableau 4 : Modélisation par équation structurelle avec la variable dépendante du refus

	Variables expliquées					
	Estimations relatives au refus	Erreurs-types	Estimations relatives à l'urbanité	Erreurs-types	Estimations relatives à la situation socioéconomique	Erreurs-types
Prédicteurs						
Taille urbaine			1,000 (0,860)fixe	0,000		
Urbanité, ville			0,856 (0,911)**	0,012		
Urbanité, code postal			0,896 (0,882)**	0,017		
Valeur de l'habitation					1,000 (0,623)fixe	0,000
Prestations sociales					-0,400 (-0,501)**	0,025
Sexe	0,070 (0,035)ns	0,024				
Âge	-0,003 (-0,046)**	0,001				
Urbanité	-0,041 (-0,057)**	0,013				
SSE	0,042 (0,054)*	0,021				
Téléphone.	0,257 (0,102)**	0,034				
Ethnicité	0,188 (0,045)**	0,058	1,324 (0,226)**	0,057	-1,715 (-0,321)**	0,082

Note : Variable dépendante avec le code 1 pour la collaboration et le code 1 pour le refus.

*p<0,05, **p<0,01; ns : non significatif.

Chi carré : 177,247 (dl= 6); EQMA : 0,048; ICA : 0,949; R² : 0,025.

Fichier de validation : Chi carré : 142,455 (dl= 5); EQMA : 0,047; ICA : 0,956; R² : 0,033.

L'ethnicité est d'un effet positif – phénomène peut-être un peu étonnant – sur les probabilités de collaboration. La population d'origine étrangère extraoccidentale a tendance à refuser de collaborer à l'enquête moins souvent que les autres unités échantillonnées. Dans l'interprétation de ces résultats, il faut user de prudence à cause des bas taux de prise de contact et des hauts taux de non-réponse imputables à l'obstacle de la langue dans le cas des minorités ethniques. Comme nous l'avons mentionné, les unités échantillonnées qui ne sont pas jointes n'ont pas la possibilité de refuser de participer à l'enquête (Hox et De Leeuw, 1998). De plus, les unités peuvent prétexter l'obstacle de la langue pour refuser amicalement de participer. Ajoutons que l'effet de l'urbanité, important pour les variables dépendantes de la réponse et de la prise de contact, importe moins dans le présent cas, d'où l'impression qu'il existe un problème de prise de contact – et non de participation – dans les régions urbaines. L'incidence de la SSE est très légère. Là encore, l'indicateur de l'existence connue du téléphone se révèle un facteur de prévision. Les unités échantillonnées plus âgées se caractérisent par des probabilités un peu supérieures de refus de participation.

4.3 Quelles variables ou prédispositions sont-elles liées au phénomène de la non-réponse, de l'absence de prise de contact et du refus dans les groupes ethniques?

Nous avons voulu savoir si les prédicteurs des divers résultats de réponse avaient la même incidence dans le cas des groupes ethniques. Nous avons donc procédé à une analyse en groupes multiples avec deux groupes une fois de plus. Le premier groupe comprenait la population de souche et la population d'origine étrangère occidentale et le second était formé de la population d'origine étrangère extraoccidentale. Nous présentons au tableau 5 les résultats de cette analyse avec la réponse comme variable de résultat.

Tableau 5 : Analyse de la réponse en groupes multiples

Prédicteurs de réponse	Variables expliquées			
	Population de souche et population d'origine étrangère occidentale		Population d'origine étrangère extraoccidentale	
	Estimations	Erreurs-types	Estimations	Erreurs-types
Sexe	0,063 (0,032)**	0,016	0,153 (0,076)**	0,058
Âge**	-0,002 (-0,022)*	0,001	-0,008 (-0,092)*	0,002
Urbanité**	-0,085 (-0,127)**	0,006	-0,180 (-0,244)**	0,032
SSE	0,021 (0,036)**	0,007	0,016 (0,027)ns	0,032
Téléphone*	0,318 (0,129)**	0,020	0,173 (0,080)*	0,074

* valeur $p < 0,05$ ** ; valeur $p < 0,01$; ns : non significatif

* la différence est significative entre les groupes pour cette variable au niveau $p < 0,05$; ** valeur $p < 0,01$.

Chi carré : 14,512 (dl=5) ; valeur p : 0,0126 ; EQMA : 0,012 ; ICA : 0,698 ; R^2_1 : 0,047 ; R^2_2 : 0,097.

Notre analyse a porté sur 26 479 membres de la population de souche et d'origine étrangère occidentale (premier groupe) et 1 893 membres de la population d'origine étrangère non occidentale (second groupe). D'après ces résultats, l'existence connue du téléphone influe plus sur les probabilités de réponse dans la population de souche et d'origine occidentale que dans la population d'origine extraoccidentale. En terme d'effets d'interaction, l'appartenance à un groupe a un effet modérateur sur la relation entre l'existence connue du téléphone et la réponse. L'incidence négative de l'urbanité sur les probabilités de réponse est seulement un peu plus marquée dans le cas de la population d'origine étrangère non occidentale. L'âge est d'un effet négatif un peu plus prononcé sur le taux de réponse dans le second groupe que dans le premier. C'est que, dans le second, le taux de non-réponse s'élève à cause de l'obstacle de la langue, phénomène presque uniquement présent dans la population d'origine étrangère non occidentale. Le tableau qui suit livre les résultats des deux groupes pour la variable dépendante de la prise de contact :

Tableau 6 : Analyse en groupes multiples de la prise de contact

Prédicteurs de la prise de contact	Variables expliquées			
	Population de souche et d'origine étrangère occidentale		Population d'origine étrangère extraoccidentale	
	Estimations	Erreurs-types	Estimations	Erreurs-types
Sexe	0,076 (0,038)**	0,020	0,152 (0,075)*	0,085
Âge*	0,006 (0,082)**	0,001	-0,002 (-0,019)ns	0,003
Urbanité**	-0,138 (-0,206)**	0,007	-0,247 (-0,336)**	0,036
SSE	0,002 (0,004)ns	0,007	0,002 (0,004)ns	0,035
Téléphone**	0,291 (0,118)**	0,022	0,198 (0,092)*	0,096

* valeur $p < 0,05$; ** valeur $p < 0,01$; ns : non significatif

Chi carré : 13,198 (dl=5) ; valeur p : 0,0215 ; EQMA : 0,011 ; ICA : 0,739 ; R^2_1 : 0,075 ; R^2_2 : 0,143.

Là aussi, l'indicateur paramétrique de la desserte téléphonique terrestre varie selon les groupes. C'est un facteur de prévision qui influe largement sur les probabilités de prise de contact dans le premier groupe. L'effet négatif de l'urbanité est par ailleurs plus prononcé dans la population d'origine étrangère extraoccidentale que dans la population de souche et d'origine étrangère occidentale. L'urbanité n'en est pas moins d'un effet négatif relativement marqué sur le taux de prise de contact dans le premier groupe (population de souche et d'origine occidentale). Enfin, le tableau 7 présente les résultats de l'analyse en groupes multiples avec le refus comme variable de résultat.

Tableau 7 : Analyse en groupes multiples du refus

Prédicteurs de refus	Variables expliquées			
	Population de souche et d'origine étrangère occidentale		Population d'origine étrangère extraoccidentale	
	Estimations	Erreurs-types	Estimations	Erreurs-types
Sexe	0,044 (0,022)*	0,018	0,142 (0,071)*	0,072
Âge*	-0,005 (-0,062)**	0,001	0,003 (0,043)ns	0,003
Urbanité*	-0,004 (-0,059)**	0,007	-0,136 (-0,179)**	0,050
SSE	0,015 (0,028)*	0,006	-0,022 (-0,039)ns	0,048
Téléphone	0,264 (0,104)**	0,022	0,185 (0,088)ns	0,096

* valeur $p < 0,05$; ** valeur $p < 0,01$; ns : non significatif

Chi carré : 2,804 (dl=2) valeur p : 0,2429; EQMA : 0,006; ICA : 0,911; R^2_1 : 0,023; R^2_2 : 0,047.

Lorsque nous avons regardé les prédicteurs de refus de participation des unités échantillonnées parmi les groupes dans cette analyse en groupes multiples, nous nous sommes uniquement attachés une fois de plus aux unités qui avaient été jointes. Nous avons ainsi obtenu 23 210 unités échantillonnées restantes pour le premier groupe et 1 359 pour le second (unités échantillonnées jointes) pour l'analyse. Dans la population d'origine étrangère non occidentale, l'urbanité a une fois de plus une incidence un peu plus négative sur le résultat de réponse, le taux de refus en l'occurrence. L'existence connue du téléphone par réseau terrestre est un bon facteur de prévision dans le premier groupe, mais ne joue guère de rôle dans le second.

5. CONCLUSIONS

Les minorités ethniques tiennent de plus en plus de place dans les sociétés occidentales et offrent donc de plus en plus d'intérêt pour les décideurs. Selon la définition de Statistique Pays-Bas, la population néerlandaise est « allochtone » dans une proportion approximative de 20 %. Selon les prévisions, la proportion de la population nationale que représentent les groupes minoritaires atteindra les 35 % en 2050 (De Jong et Hilderink, 2004). On éprouve sans cesse plus le besoin de recueillir des données sur les minorités ethniques, mais des doutes subsistent quant à la qualité de cette information, les groupes ethniques présentant des taux de non-réponse supérieurs. En soi, la non-réponse ne veut pas automatiquement dire que les estimations ponctuelles sont entachées d'un biais. Les taux de non-réponse n'en sont pas moins l'indice de problèmes de biais possibles. Dans la lutte livrée à la non-réponse, on devrait surtout s'employer à réduire l'erreur de non-réponse. En tentant simplement d'accroître les taux de réponse, on peut en réalité se trouver à augmenter aussi l'erreur d'enquête (Merkle et Edelman, 2002), aussi importe-t-il de repérer la non-réponse dans les groupes sociaux de manière à pouvoir adopter des stratégies propres à abaisser les taux de non-réponse dans les sous-groupes qui sont précisément sous-représentés.

Les analyses du présent article reposent sur les résultats de l'enquête néerlandaise de 1998 sur les conditions de vie. Les tableaux à deux variables de la réponse et de l'ethnicité font voir de grandes différences de taux de réponse entre groupes ethniques. Une de nos constatations d'intérêt est l'importance des taux d'absence de prise de contact dans les minorités ethniques et, plus particulièrement, dans la population d'origine étrangère extraoccidentale. L'explication ne peut en être ramenée aux seuls problèmes qui se sont posés sur le terrain cette année-là en raison de l'insuffisance des capacités d'interview. Depuis, beaucoup de choses ont changé à Statistique Pays-Bas et, de ce fait, les taux de réponse se sont améliorés. Pour élever les taux de réponse, on a normalisé les stratégies de prise de contact et de persuasion dans l'organisme (Snijkers et Kockelkoren, 2004). Fait remarquable, les taux de collaboration se sont plus élevés dans les minorités ethniques que dans la population de souche. Il reste que, si les unités échantillonnées ne sont pas jointes, il leur est, bien sûr, impossible de refuser de collaborer à une enquête (Hox et De Leeuw, 1998). À cet égard, un relèvement du minimum de tentatives de prise de contact après un premier échec à Statistique Pays-Bas en mars 2004 a eu pour effet d'accroître nettement les taux de prise de contact et de réponse de la population d'origine étrangère non occidentale, mais non pas ses taux de refus (Schmeets, 2005).

Les résultats de modélisation par équation structurelle montrent que l'effet négatif de l'ethnicité sur la réponse disparaît en partie une fois d'autres variables prises en compte. Notons en particulier que l'urbanité influe largement sur la non-réponse, mais que l'ethnicité a toujours une incidence sur les probabilités de réponse. Sachant que, dans le cas des minorités ethniques, la non-réponse existe surtout dans les régions hautement urbanisées, le chercheur peut se concentrer sur cette partie bien précise de la population. Bien sûr, la corrélation est étroite entre milieu urbain et

minorités ethniques, mais on constate, aux Pays-Bas par exemple, que, dans une proportion approximative de 70 %, la population allochtone (et 70 % de la population d'origine étrangère extraoccidentale) n'habite pas une des quatre grandes villes du pays. Depuis longtemps, on a établi un lien entre l'urbanité et la non-réponse aux enquêtes (Groves et Couper, 1998). Si les taux de non-réponse sont plus élevés aux Pays-Bas que dans la plupart des autres pays, c'est principalement en fait que la partie ouest de notre pays est particulièrement urbanisée. La réponse demeure toutefois un processus qui loge nettement à l'enseigne de la chance. Il y a dans la société des groupes comme les minorités ethniques qui présentent bel et bien des taux de réponse inférieurs. On se rend compte heureusement que, même en se munissant d'une riche information de contexte, il reste extrêmement difficile de prévoir si quelqu'un répondra ou non à une enquête. Bien que disposant d'une riche information administrative, on peut voir par un bas pseudo- R^2 de 0,072 que la non-réponse n'a guère de caractère systématique. Même avec des coefficients de régression et des R^2 de faibles valeurs, on devrait néanmoins s'employer à accroître la réponse dans les régions urbaines, surtout là où se posent des problèmes de prise de contact. En adaptant les stratégies de collecte de données afin d'obtenir de meilleurs taux de prise de contact, on se trouverait aussi à augmenter les taux de réponse des minorités ethniques.

Les résultats de l'analyse en groupes multiples indiquent que les groupes ethniques ne sont pas homogènes dans leurs mécanismes de réponse. L'âge et l'urbanité sont d'un plus grand effet négatif sur les probabilités de réponse et de prise de contact dans la population d'origine étrangère extraoccidentale que dans toute la population de souche et d'origine étrangère occidentale. Par ailleurs, l'indicateur de l'existence connue du téléphone par desserte terrestre se révèle un très bon facteur de prévision de la réponse dans le cas des Néerlandais de souche et d'origine étrangère occidentale. L'effet négatif est bien moins marqué dans la population d'origine étrangère non occidentale. Bien sûr, les résultats de notre étude ne sont que partiellement généralisables à l'échelle des autres pays, émanant d'une enquête menée en sol néerlandais. Notre analyse n'en indique pas moins que, dans une situation d'infériorité des taux de réponse des minorités ethniques, on devrait plutôt s'efforcer d'accroître la réponse dans les régions urbaines. On devrait redoubler d'efforts pour élever les taux de prise de contact en milieu urbain. Une solution possible serait d'aborder les unités échantillonnées par des stratégies sur mesure où, par exemple, le minimum de tentatives de prise de contact serait plus haut que dans le cas des unités échantillonnées en milieu non urbain.

RÉFÉRENCES

- The American Association for Public Opinion Research (2000), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, Ann Arbor, Michigan: AAPOR.
- Feskens, R.C.W. (2005), "Het Afbakenen en Benaderen van Allochtonen in internationaal Perspectief", in Schmeets, H. and Bie, van der, R. (eds.) *Over het omgaan met Allochtonen in Enquêteonderzoek*, Heerlen: Centraal Bureau voor de Statistiek (en cours).
- Groves, R.M. (1989), *Survey Errors and Survey Costs*, New York: Wiley.
- Groves, R.M. et Couper, M.P. (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley.
- De Heer, W. (1999), "International Response Trends: Results of an International Survey", *Journal of Official Statistics*, 15, pp. 129-142.
- De Heer, W. et De Leeuw, E. (2002), "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison", in Groves R. et al. (eds.) *Survey Nonresponse*, New York: John Wiley, pp. 41-54.
- Hox, J.J. et De Leeuw, E. (1998), "Nonresponse in Surveys: Een Overzicht", *Kwantitatieve Methoden*, 19, pp. 31-53.
- De Jong, A. et Hilderink, H. (2004), "Lange termijn allochtonenscenario's voor Nederland", *Bevolkingstrends*, 52 (1), Voorburg/Heerlen: Centraal Bureau voor de Statistiek, pp. 77-82.
- Kline, R.B. (1998), *Principles and Practice of Structural Equation Modeling*, New York: Guilford Press.

- Merkle, D.M. et Edelman, M. (2002), " Nonresponse in Exit Poll: A Comprehensive Analysis", in Groves R. et al. (eds.) *Survey Nonresponse*, New York: John Wiley, pp. 243-257.
- Muthén, B.O. (1998-2004), *Mplus Technical Appendices*, Los Angeles, CA: Muthén & Muthén
- Muthén, L.K. Muthén, B.O. (1998-2004), *Mplus Users's Guide*, Third Edition. Los Angeles, CA: Muthén & Muthén.
- Reep, C. (2003), "Moeilijk Waarneembare Groepen. Een inventarisatie", CBS-report H1568-03-SOO, Heerlen: Centraal Bureau voor de Statistiek.
- Schmeets, H. et Michiels, J. (2003), "Het Effect van Non-respons onder Allochtonen", *Bevolkingstrends*, 51 (4), Voorburg/Heerlen: Centraal Bureau voor de Statistiek, pp. 52-57.
- Schmeets, H., Reep, C. et Snijkers, G. (2003), "Afbakenen van moeilijk waarneembare bevolkingsgroepen", *Bevolkingstrends*, 51 (3), Voorburg/Heerlen: Centraal Bureau voor de Statistiek, pp. 62-68.
- Schmeets, H. (2005), "De invloed van het aantal contactpogingen op de respons van allochtonen", in Schmeets, H. and Bie, van der, R. (eds.) *Over het omgaan met Allochtonen in Enquêteonderzoek*, Heerlen: Centraal Bureau voor de Statistiek (en cours).
- Schouten, B. (2003), "De verschillende Typen Non-respons in POLS 1998", CBS-report 2319-03-TMO. Heerlen: Centraal Bureau voor de Statistiek.
- Snijkers, G. (2003), "Programmaplan Strategisch Onderzoek 'Non-respons en Moeilijk Waarneembare Groepen': Naar een optimale Meetstrategie", CBS-report H3318-03-SOO, Heerlen: Centraal Bureau voor de Statistiek.
- Snijkers, G. et Kockelkoren. S. (2004), "De gestandaardiseerde Benaderingsstrategie 2003: Evaluatie en Aanbevelingen", CBS-report H415-04-SOO, Heerlen: Centraal Bureau voor de Statistiek.
- Stoop, I.A.L. (2004), "Surveying Nonrespondents", *Field Methods*, Vol. 16, N° 1, février 2004, pp. 23-54.
- Thornberry, O.T. ,Jr., et. Massey, J.T. (1988), "Trends in United States Telephone Coverage across Time and Subgroups", in Groves R. et al. (eds.) *Telephone Survey Methodology*, New York: John Wiley & Sons, pp: 25-49.

ANNEXE

Variables :

Sexe; sexe de l'unité échantillonnée :

- 0 : homme
- 1 : femme

Âge; âge de l'unité échantillonnée :

- 15-65 ans

Taille urbaine; taille de la collectivité où l'unité échantillonnée est inscrite :

- 1 : petite; 8 : grande

Urbanité de la collectivité; urbanité au niveau de la collectivité où l'unité échantillonnée est inscrite :

- 1 urbanisée
- 2 à peine urbanisée
- 3 légèrement urbanisée
- 4 urbanisée
- 5 hautement urbanisée

Urbanité au niveau des codes postaux; urbanité au niveau du code postal pour l'adresse où l'unité échantillonnée est inscrite :

- 1 moins de 500 adresses au km²
- 2 de 500 à 1 000 adresses au km²
- 3 de 1 000 à 1 500 adresses au km²
- 4 de 1 500 à 2 500 adresses au km²
- 5 plus de 2 500 adresses au km²

Indicateurs de desserte téléphonique terrestre connue (service du téléphone); existence connue du téléphone pour l'unité échantillonnée :

- 0 : sans téléphone connu
- 1 : téléphone connu

Réponse; réponse (partielle) de l'unité échantillonnée :

- 0 : non-réponse
- 1 : réponse

Valeur de l'habitation; valeur de l'habitation (en florins néerlandais) où l'unité échantillonnée est inscrite :

- 1 moins de 50 mille
- 2 50 – 75 mille
- 3 75 – 100 mille
- 4 100 – 125 mille
- 5 125 – 150 mille
- 6 150 – 200 mille
- 7 200 – 250 mille
- 8 250 – 300 mille
- 9 300 – 350 mille
- 10 350 – 400 mille
- 11 400 – 500 mille
- 12 plus de 500 mille

Indicateur de réception quelconque de prestations sociales; réception quelconque de prestations sociales par l'unité échantillonnée :

- 0 : non
- 1 : oui

Ethnicité; unité échantillonnée d'origine étrangère extraoccidentale :

0 : non

1 : oui

Prise de contact; contact établi avec l'unité échantillonnée pendant les opérations de terrain :

0 : non (sans prise de contact)

1 : oui (prise de contact)

Refus; refus de participation à l'enquête par l'unité échantillonnée :

0 : refus de fournir l'information demandée de la part de l'unité échantillonnée

1 : collaboration de l'unité échantillonnée