

## QUALITÉ DES DONNÉES DES QUESTIONS SUR LE SALAIRE POUR L'ENQUÊTE CANADIENNE SUR LA POPULATION ACTIVE

François Brisebois<sup>1</sup> et Charlene Walker<sup>1</sup>

### RÉSUMÉ

Suite au recensement de la population canadienne de 1991, l'Enquête canadienne sur la population active a été soumise à un cinquième remaniement depuis son introduction en 1952. Au nombre des changements apportés lors de ce dernier remaniement, figurait la mise en place d'un nouveau questionnaire visant à mieux refléter les réalités du marché du travail d'aujourd'hui. Le nouveau questionnaire, introduit en janvier 1997, inclut maintenant plus de questions portant sur l'emploi, dont entre autres, des questions portant sur le revenu. Étant donné la nature plutôt délicate du sujet, ces questions sont souvent sujettes à des taux d'erreurs non dues à l'échantillonnage (plus spécifiquement des erreurs de réponse et erreurs dues à la non-réponse) plus élevés que ceux associés à d'autres types de questions. Cette étude tente d'évaluer l'ampleur de l'erreur non due à l'échantillonnage présente dans les questions portant sur le salaire à l'aide de certains indicateurs de qualité.

MOTS-CLÉS: Qualité des données; erreurs de réponse; erreurs dues à la non-réponse; salaire.

### ABSTRACT

Following the 1991 Canadian Census of Population, the Canadian Labour Force Survey underwent its fifth redesign since its introduction in 1952. Among the changes in this redesign, was the implementation of a new questionnaire intended to better reflect today's labour market. The new questionnaire, implemented in January 1997, includes more questions about employment, including among others, questions about earnings. Due to the sensitivity of the topic, these questions are often subject to higher rates of non-sampling error (specifically response and non-response error) than less sensitive questions. This paper will assess the extent of the non-sampling error to the new earnings questions with some quality indicators.

KEY WORDS: Data quality; Response errors; Non-response errors; Earnings.

### 1. INTRODUCTION

Au cours des dernières années, le marché du travail a changé de façon radicale : une entrée massive des femmes, les bébés du baby boom ont atteint l'âge mûr; les marchés mondiaux ont changé l'économie; puis les employeurs se sont tournés vers des stratégies de travail flexibles afin de faire face à l'incertitude et à la fluctuation de la demande des consommateurs. Par conséquent, après vingt ans avec le même questionnaire, l'Enquête canadienne sur la population active (EPA) a effectué un remaniement de son questionnaire. Un des changements apportés suite à la demande générale des principaux utilisateurs, de différents ministères fédéraux et provinciaux, des institutions scolaires et de différents groupes d'affaires, a été l'ajout de questions portant sur le

salaire. Le nouveau questionnaire a été introduit de façon progressive de septembre 1996 à janvier 1997.

Les questions sur le salaire ne sont pas choses nouvelles dans le monde des enquêtes. Plusieurs des enquêtes de Statistique Canada incluent de telles questions. Aux États-Unis, le 'Current Population Survey' inclut des questions sur le salaire depuis plus de 15 ans. En 1994, ils ont remanié leur questionnaire afin d'offrir une variété de choix de réponses (par exemple, salaire horaire, hebdomadaire, mensuel ou annuel). Leur expérience a été très positive et la non-réponse aux questions sur le salaire était faible. Au Canada, une série de questions semblables a été adoptée au cours du remaniement du questionnaire de l'EPA. Ces questions ont été minutieusement testées et une formation complète a été offerte aux

<sup>1</sup> François Brisebois et Charlene Walker, Statistique Canada, Édifice R.H. Coats, 16<sup>e</sup> étage, Ottawa, Ontario, Canada, K1A 0T6, [brisfra@statcan.ca](mailto:brisfra@statcan.ca) et [budzicd@statcan.ca](mailto:budzicd@statcan.ca).

intervieweurs avant d'introduire le nouveau questionnaire.

## 2. LES QUESTIONS SUR LE SALAIRE À L'EPA

Le questionnaire de l'EPA est constitué de deux composantes; la composante démographique et celle du questionnaire de l'EPA. Les questions de la composante démographique sont demandées à tous les membres du ménage alors que celles du questionnaire de l'EPA sont demandées aux personnes de 15 ans et plus. La composante du questionnaire de l'EPA inclut des questions portant sur une grande variété de sujets reliés à l'emploi, tels que le type d'emploi, l'entreprise, la syndicalisation, les heures travaillées et le salaire.

Toutes les questions concernant l'emploi principal sont demandées à l'intérieur d'un même bloc. Les questions relatives au salaire ne sont demandées que pour l'emploi principal et sont situées vers la fin du bloc. Elles ne sont aussi posées que lors de l'interview initiale aux personnes étant employées et recevant un salaire ou un traitement de leur employeur. Pour réduire le fardeau de réponse, les questions concernant le salaire sont seulement demandées lors des interviews subséquentes s'il y a eu un changement dans la situation d'emploi depuis le dernier interview. S'il n'y a pas eu de changement, l'information obtenue lors de l'interview initiale est automatiquement reportée aux mois subséquents.

Pour mesurer le salaire, l'EPA utilise une série de questions qui cherche à trouver la façon la plus simple et exacte de rapporter le salaire pour chaque individu. La première question, concernant le salaire, demande '*... est-[il/elle] payé(e) à l'heure?*'. Dans l'affirmative, les deux prochaines questions sont '*... reçoit-[il/elle] habituellement des pourboires ou des commissions?*' et '*Quel est le salaire horaire de ...?*'. Toutefois, si la réponse à la première question est non, les deux questions qui suivent sont plutôt '*Quelle est la manière la plus facile de nous dire le salaire ou le traitement de ..., avant impôt et autres déductions? Sous forme de gains annuels, mensuels, hebdomadaires ou autres?*' et '*Quel est le salaire ou le traitement [annuel/mensuel/bimensuel/aux deux semaines/ hebdomadaire] de ...?*'. Des règles de contrôle pour les valeurs maximales et minimales des salaires sont implantées dans les ordinateurs portatifs des intervieweurs et dépendent évidemment de la période utilisée pour rapporter le salaire.

À cause du caractère délicat des questions sur le salaire et de la difficulté de répondre de façon précise

à ces questions, la présence d'erreurs de réponse et d'erreurs dues à la non-réponse est inévitable et aussi généralement difficile à mesurer. Toutefois, avec l'implantation de la technologie d'interview assistée par ordinateur, de nouvelles mesures maintenant disponibles peuvent être utilisées pour obtenir certains indicateurs de qualité et ainsi évaluer l'ampleur et l'effet des erreurs de réponse et dues à la non-réponse.

## 3. ERREURS DE RÉPONSE ET ERREURS DUES À LA NON-RÉPONSE

Les *erreurs de réponse* sont définies comme étant des réponses inexactes, de piètre qualité et peuvent être causées soit par l'intervieweur (en se faisant mal comprendre) ou par le répondant (ne comprend pas la question, erreur de rappel, inquiétudes relatives à la vie privée, etc.). Les erreurs de réponse peuvent introduire un biais si elles prennent systématiquement toutes une même direction ou peuvent faire augmenter la variance si elles sont plutôt aléatoires.

Les *erreurs dues à la non-réponse* sont définies par l'absence de réponse à une question, absence due soit à une non-réponse complète de l'unité échantillonnée ou soit à une non-réponse à une question seulement. Cette étude s'intéresse plus particulièrement aux cas de non-réponse au niveau des questions sur le salaire. Les erreurs dues à la non-réponse augmentent la variance puisque la taille d'échantillon se voit être réduite et introduisent un biais si la non-réponse n'est pas aléatoire et qu'elle est corrélée avec d'autres variables mesurées par l'enquête. Il est préférable de prévenir la non-réponse plutôt qu'essayer de compenser pour celle-ci après coup. Pour ce faire, l'EPA réduit le fardeau de réponse en demandant les questions sur le salaire que lors de l'interview initiale, en justifiant son besoin auprès du répondant et en assurant la confidentialité.

Cette étude évalue l'ampleur des erreurs de réponse et celles dues à la non-réponse pour les questions sur le salaire en analysant certains indicateurs de qualité. Les deux prochaines sections décrivent les indicateurs analysés et présentent les résultats obtenus avec les données mensuelles de 1997.

### 3.1 Erreurs de réponse

Les erreurs de réponse examinées dans cette étude concernent l'exactitude des données recueillies sur les salaires. Parce qu'il est parfois difficile pour un répondant de fournir un montant exact pour le salaire, une approximation peut s'avérer un bon compromis à la non-réponse. Ceci est particulièrement vrai dans le cas des réponses obtenues par procuration alors que

des membres du ménage sont absents. Lorsque l'intervieweur demande le montant du salaire au répondant, il note (dans un champ réservé à cette fin sur l'ordinateur portable) tout commentaire fait relatif à l'exactitude du salaire donné. Ainsi, si le répondant mentionne que le montant donné est approximatif, ceci sera indiqué au dossier du répondant. L'utilisation d'approximations est acceptable mais peut toutefois causer un certain biais; les gens auront peut-être tendance à donner des montants plus élevés ou plus faibles que les montants réels. C'est ce que l'on cherche à vérifier.

Pour ce faire, les notes recueillies pour la question demandant le salaire ont d'abord été codées en un certain nombre de catégories. Les résultats de ce codage ont montré que 51,2% des notes mentionnaient sous une certaine forme que le montant rapporté était approximatif. Par la suite, deux groupes de personnes ont été formés; un incluant les personnes pour lesquelles une note au dossier indiquait que le salaire déclaré était approximatif, puis un autre incluant celles n'ayant aucune note relative au salaire. L'hypothèse est que si les réponses ne sont pas biaisées, on devrait en moyenne avoir le même salaire pour les deux groupes. Il est vrai que cette comparaison est peut-être un peu grossière parce qu'elle ne tient pas compte d'autres variables pouvant influencer, telles que la strate de revenus, l'âge, etc.. Elle trace toutefois un bon portrait global de la situation.

Les résultats présentés au Tableau 1 montrent qu'au Canada, la moyenne des salaires des deux groupes est très semblable: \$15,83/hr pour les 'approximations' vs \$15,60/hr pour les 'exacts'. Ceci laisse donc croire qu'il n'y a pas vraiment de biais causé par l'acceptation d'un montant approximatif dans le cas où le montant exact n'est pas connu par le répondant. Toutefois, lorsqu'on observe les moyennes obtenues pour les différentes provinces, on remarque qu'en général, le groupe des 'approximations' présente des moyennes plus élevées que le groupe des 'exacts' dans les provinces où les salaires moyens (exacts) sont les plus bas, i.e. pour les provinces des maritimes. En Colombie-Britannique, province ayant le salaire horaire moyen le plus élevé, on observe la tendance inverse alors que la moyenne des 'approximations' est

d'environ \$1,30/hr plus basse que celle des 'exacts'.

### 3.2 Erreurs dues à la non-réponse

Deux indicateurs de qualité relatifs à l'erreur de non-réponse sont analysés dans cette section, il s'agit des refus de répondre et des cas de 'ne sait pas'. Comme il a été mentionné auparavant, les questions sur les salaires sont de nature plus personnelle et donc plus sujettes à des refus de la part des personnes interviewées. Aussi, la nature exacte des questions fait qu'elles sont plus difficiles à répondre, résultant en un nombre plus élevé de 'ne sait pas'.

Pour vérifier la première affirmation, les taux de refus pour chaque question du bloc portant sur les salaires ont été calculés au niveau national et sont présentés dans le Tableau 2. Le tableau inclut aussi les taux de refus pour d'autres questions étant considérées elles aussi à un certain point, de nature personnelle. Depuis le dernier remaniement, l'EPA inclut une étape supplémentaire de stratification basée sur les revenus, ceci afin d'avoir une bonne représentativité de la population et ainsi d'avoir des résultats plus exacts pour les estimations de salaires. Environ 1% des ménages interviewés font partie des strates à revenus élevés. Le Tableau 2 inclut les taux de refus obtenus pour les ménages faisant partie des strates à revenus élevés.

Comme on pouvait s'y en attendre, les questions portant sur les salaires ont des taux de refus plus élevés que les autres questions, particulièrement la question demandant le montant du salaire proprement dit. Le taux pour cette question est environ dix fois plus élevé que les taux observés pour certaines autres questions aussi perçues comme étant de nature personnelle, telles que la date de naissance, la syndicalisation et le fait d'être couvert ou non par une convention collective. En ce qui concerne les résultats dans les strates à revenus élevés, la question demandant le montant du salaire montre un taux de 5,0%, comparativement à 2,9% pour l'ensemble de l'échantillon. Les taux pour les autres questions référant au salaire présentent tous des taux assez semblables à ceux observés pour l'ensemble de l'échantillon.

**Tableau 1: Comparaison des salaires horaires moyens pour les personnes déclarant des montants 'approximatifs' vs celles déclarant des montants 'exacts', Canada et par province (en \$).**

Groupes/Prov.	Canada	T.-N.	I.-P.-É.	N.-É.	N.-B.	Qué.	Ont.	Man.	Sask.	Alberta	C.-B.
Exacts	15,60	13,23	11,82	12,90	12,99	15,37	16,32	14,92	13,35	14,52	16,96
Approximatifs	15,83	14,75	14,06	14,21	13,53	15,55	16,85	13,81	13,65	14,77	15,68

**Tableau 2: Taux de refus pour les questions sur les salaires et pour certaines questions de nature personnelle, global et pour les strates à revenus élevés**

Question	Taux global	Strates à revenus élevés
Payé à l'heure?	0,3%	0,3%
Reçoit des commissions et/ou pourboires?	0,2%	0,2%
Manière la plus facile de dire le salaire?	1,7%	1,6%
Quel est le salaire?	2,9%	5,0%
Date de naissance?	0,3%	1,6%
Est syndiqué?	0,2%	0,0%
Couvert par une convention collective?	0,2%	0,0%

Au niveau provincial, les taux observés diffèrent considérablement d'une province à l'autre, du moins en ce qui concerne la question demandant le montant du salaire. Le Tableau 3 présente les taux pour cette question de même que les salaires horaires moyens pour chacune des dix provinces canadiennes. Le Tableau 2 donnait une indication que les taux de refus pouvaient être influencés par la magnitude du salaire. On observe cette même indication dans le Tableau 3 alors que les provinces ayant les taux de refus les plus élevés sont celles ayant les salaires horaires les plus élevés. En effet, les provinces de l'ouest canadien, i.e. le Manitoba, l'Alberta, la Saskatchewan et la Colombie-Britannique, montrent tous des taux de refus autour de 4,0% et des salaires horaires plus élevés que les provinces des maritimes, où l'on observe à l'inverse des taux de refus très faibles.

Ces derniers résultats sembleraient nous laisser croire que le taux de refus à la question demandant le salaire est en grande partie relié à la magnitude du salaire. Il faut toutefois aussi considérer d'autres facteurs pouvant influencer le taux de refus, comme l'attitude des gens d'une province à l'autre et les procédures opérationnelles utilisées dans les différents bureaux régionaux de Statistique Canada. Malgré le fait que les principes généraux entourant le processus d'interview soit standard d'un bureau à l'autre, certains aspects peuvent être différents, comme par

exemple la formation, la charge de travail, le temps perdu pour de plus longs déplacements, etc.

Autre résultat intéressant en ce qui concerne les refus aux questions sur les salaires; parmi tous les répondants à l'enquête, seulement 2,6% ont refusé de répondre à au moins une question. Ce qui est encore plus marquant est que parmi ces 2,6%, environ 80% ont refusé seulement à des questions du bloc portant sur le salaire. Un refus aux questions sur le salaire ne semble donc pas compromettre le reste de l'interview.

Pour tenter de tracer un portrait des personnes refusant de dévoiler leur salaire, le profil des personnes ayant refusé de répondre à la question demandant le salaire a été comparé à celui des personnes l'ayant répondu. Un point important doit être souligné ici; seuls les gens ayant répondu pour eux-mêmes ont été conservés pour cette comparaison, on exclut donc toute réponse par procuration. Les deux groupes, les répondants vs les refus, ont été comparés selon les variables démographiques suivantes : le sexe, l'âge, l'éducation et l'état marital. Parmi ces variables, seul l'âge s'est avéré être différent de façon significative entre les deux groupes comparés; l'âge moyen des personnes répondantes est de 35,8 ans et celle des personnes ayant refusé est de 40,3 ans. Le Tableau 4 résume les résultats obtenus pour ces comparaisons.

**Tableau 3: Taux de refus et salaire horaire moyen pour chacune des dix provinces canadiennes**

Province	Taux de refus	Salaire horaire moyen
Terre-Neuve	1,2%	\$13,24
Île-du-Prince-Édouard	1,0%	\$11,82
Nouvelle-Écosse	2,3%	\$12,91
Nouveau-Brunswick	0,9%	\$12,99
Québec	3,3%	\$15,37
Ontario	2,1%	\$16,32
Manitoba	4,0%	\$13,81
Saskatchewan	4,2%	\$13,64
Alberta	4,5%	\$14,77
Colombie-Britannique	3,8%	\$16,95

**Tableau 4: Comparaison des répondants à ceux ayant refusé de répondre à la question demandant le salaire**

Variables démographiques comparées	Profil des groupes		Taux de refus
	Répondants	Refus	
Sexe			
<i>Femmes</i>	58,2%	53,8%	2,2%
<i>Hommes</i>	41,8%	46,2%	2,6%
Âge	35,8 ans	40,3 ans	
Éducation			
<i>Aucun diplôme ou cert. d'études post-secondaires</i>	14,7%	12,1%	1,8%
<i>Diplôme ou cert. de métier</i>	20,6%	27,3%	2,9%
<i>Diplôme ou cert. non-universitaire</i>	32,2%	29,0%	2,0%
<i>Cert. universitaire inférieur au baccalauréat</i>	4,7%	4,9%	2,3%
<i>Diplôme ou certificat universitaire</i>	27,8%	26,7%	2,1%
État marital			
<i>Marié(e) ou vivant en union libre</i>	68,3%	66,0%	2,4%
<i>Célibataire (jamais marié(e))</i>	17,2%	21,4%	1,9%
<i>Veuve ou veuf</i>	1,9%	1,6%	2,7%
<i>Séparé(e) ou divorcé(e)</i>	12,6%	11,0%	2,7%

En plus des refus, il existe une deuxième forme de non-réponse qui s'avère elle aussi inévitable dans tout enquête, il s'agit des cas de 'ne sait pas'. On réfère ici aux cas où les personnes interviewées ne connaissent pas la réponse à la question posée. Il faut se rappeler que l'EPA permet la réponse par procuration, ce qui aura l'effet d'augmenter le nombre de 'ne sait pas'. Toutefois, dans le cas des questions sur les salaires, une réponse approximative est acceptée lorsque le répondant ne connaît pas exactement le montant à déclarer. Ceci devrait avoir l'effet inverse de réduire le nombre de 'ne sait pas'. Le Tableau 5 présente les taux de 'ne sait pas' pour l'ensemble de l'échantillon, puis aussi par type de répondant, i.e. si la réponse était obtenue de la personne elle-même ou par une personne interposée. Il est important de préciser que les procédures d'interviews varient quelque peu dans les cas où un 'ne sait pas' est obtenu pour la première question du bloc sur les salaires, "... *est-til/elle] payé(e) à l'heure*". Dans ces cas, l'intervieweur poursuivra avec la question demandant "... *la manière la plus facile de nous dire le salaire ...*". Cette procédure permet de réduire le nombre de 'ne sait pas' à la question du salaire pour les cas où le répondant ne sait pas nécessairement si le salaire est payé à l'heure ou non, mais a une idée du salaire mensuel ou annuel.

En ce qui concerne les résultats obtenues, on observe

d'abord que les taux sont relativement faibles pour la question demandant si la personne est payée à l'heure, de même que pour la question suivante concernant les commissions et pourboires. Les deux autres questions montrent toutefois des taux assez élevés, atteignant presque 4% pour la question qui nous intéresse plus particulièrement, celle demandant le salaire. Les taux ont été décomposés par type de répondant et ceci apporte une bonne explication au taux élevé obtenu pour la question sur le salaire. Dans les cas de réponse par procuration, on obtient un taux de 'ne sait pas' de 6,52%, alors qu'on observe un taux de seulement 1,09% pour les cas en personne. Ce dernier taux peut sembler étrange car il représente des cas où les personnes ne connaissent pas leur propre salaire. Dans le cas où la personne ne connaîtrait pas le montant exact de son salaire, elle est encouragée à au moins donner un montant approximatif. Après une telle démarche, l'obtention de cas de 'ne sait pas' peut donc être perçue comme une forme de refus déguisé, la personne voulant ainsi éviter de répondre et d'avoir à justifier ou argumenter un refus. Il demeure que la réponse par procuration est une procédure qui facilite le processus d'interview et ne semble donc pas, selon les résultats du Tableau 5, causer de problème majeur à l'obtention des réponses pour les questions sur le salaire.

Tableau 5: Taux de "ne sait pas" pour les questions sur les salaires, total et par type de réponse

Question	Taux global	Type de répondant	
		En personne	Par procuration
Payé à l'heure?	0,68%	0,17%	1,22%
Reçoit des commissions et/ou pourboires?	0,32%	0,10%	0,56%
Manière la plus facile de dire le salaire?	4,84%	1,07%	8,92%
Quel est le salaire?	3,73%	1,09%	6,52%

#### 4. CONCLUSION

L'analyse des données portant sur les salaires a permis d'obtenir plusieurs résultats intéressants. La comparaison des salaires moyens entre réponses exactes et réponses approximatives a permis de vérifier que le fait d'accepter des réponses approximatives ne cause pas de biais pour les estimations au niveau national. L'analyse des taux de refus a montré que malgré la nature personnelle du sujet, l'EPA réussit bien avec les questions sur les salaires. Des différences de taux entre les provinces et selon l'âge des répondants ont été observées. En ce qui concerne les cas de 'ne sait pas', les taux obtenus étaient très encourageants, étant presque nuls dans le cas de réponse en personne et relativement faibles pour les cas de réponse par procuration. Ces premières analyses exploratoires se sont donc avérées positives et encouragent la poursuite des études sur la qualité des données des questions sur le salaire.

#### REMERCIEMENTS

Les auteurs remercient John Rowland, plus particulièrement pour ces nombreuses explications sur le fonctionnement des différents systèmes de traitement des données utilisés par l'EPA. Des remerciements aussi à Johanne Tremblay et Frank Mayda pour leurs précieux commentaires.

#### RÉFÉRENCES

- Walker, C., Brown, A. and Veevers, R. (1996). Data validation in the CAI world. *Proceedings of the Effects of Changing Data Collection Mode Section*, American Statistical Association, 1996.
- Statistique Canada (1997). *Bulletin : Redesigned LFS Questionnaire (Wage questions)*. Document interne.