

No 11-522-XIF au catalogue

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

**Symposium 2006 : Enjeux
méthodologiques reliés à la
mesure de la santé des
populations**



2006



**Statistics
Canada**

**Statistique
Canada**

Canada

L'effet de la spécification du modèle sur les données résultant de l'imputation multiple : leçons tirées du projet DC-HOPE

Marie G. Gantz et M. Nabil El-Khorazaty, pour l'initiative NIH-DC¹

Résumé

Le projet Healthy Outcomes of Pregnancy Education du district de Columbia est un essai randomisé financé par le National Institute of Child Health and Human Development et visant à mettre à l'essai l'efficacité d'une intervention intégrée de sensibilisation et de counselling (IISC) comparé aux soins habituels (SH) afin de réduire quatre comportements à risque chez les femmes enceintes. Les participantes ont été interviewées quatre fois. On a eu recours à la technique de l'imputation multiple pour estimer les données des interviews manquantes. La technique de l'imputation multiple a été appliquée deux fois : toutes les données ont été imputées simultanément une première fois et les données concernant les femmes des groupes IISC et SH ont été imputées séparément une deuxième fois. Les résultats des analyses des ensembles de données imputées et des données avant imputation sont comparés.

MOTS-CLÉS : Imputation multiple, principe de vouloir traiter, essai randomisé, mesures répétées.

1. Introduction

Le projet Healthy Outcomes of Pregnancy Education du district de Columbia (DC-HOPE, initiative d'information sur la grossesse) s'inscrit dans une initiative des National Institutes of Health (NIH, instituts nationaux de la santé) visant à réduire le taux de mortalité infantile chez les groupes démographiques minoritaires, à Washington, DC. Financé par le National Institute of Child Health and Human Development (NICHD, institut national de la santé de l'enfant et du développement humain), le projet DC-HOPE est un essai randomisé visant à évaluer l'efficacité d'une intervention intégrée de sensibilisation et de counselling afin de réduire quatre comportements à risque chez les femmes enceintes. Les femmes étaient admissibles au projet si elles avaient elles-mêmes déclaré être afro-américaines ou latino-américaines. De plus, elles devaient avoir au moins 18 ans et parler l'anglais, et le fœtus devait avoir un âge gestationnel estimatif (AGE) de moins de 29 semaines au moment de l'inscription. Les participantes ont été recrutées dans six cliniques du district de Columbia. Les femmes à la recherche de soins prénataux dans les cliniques ont été priées de répondre à de brèves questions de présélection informatisées (Auto-entretien assisté par ordinateur portable avec interface audio) (voir Thornberry et coll., 2003). Les femmes admissibles qui ont obtenu des résultats positifs relativement aux risques que sont le tabagisme, l'exposition à la fumée de tabac ambiante (EFTA), la dépression ou la violence du partenaire intime (VPI) ont été invitées à prendre part à l'étude. Des participantes ont été choisies au hasard pour assister à des séances d'intervention intégrée de sensibilisation et de counselling (IISC), qui coïncidaient avec les soins prénataux, ou pour recevoir les soins habituels (SH) offerts aux femmes enceintes.

Entre juillet 2001 et octobre 2003, 2 913 femmes ont rempli le questionnaire de présélection et 1 070 d'entre elles ont été inscrites au projet DC-HOPE. La vaste majorité des femmes inscrites ont elles-mêmes déclarées être des Afro-américaines, et les 22 femmes qui ne l'ont pas fait ont été exclues de l'analyse finale. En outre, quatre femmes ont été jugées inadmissibles parce qu'elles n'avaient pas répondu aux questions de base requises avant la fin de leur grossesse. Sur les 1 044 femmes retenues pour l'analyse finale, 521 ont été aléatoirement affectées au groupe IISC et 523, au groupe SH.

¹ Marie G. Gantz, Ph.D., RTI International, 3040, Cornwallis Road, Research Triangle Park, Caroline du Nord, États-Unis, 27712; M. Nabil El-Khorazaty, Ph.D., RTI International, 6110, Executive Boulevard, bureau 902, Rockville, Maryland, États-Unis, 20852. Le projet NIH-DC est une initiative conjointe du Children's National Medical Center, du SPHHS de l'Université George Washington, du centre médical de l'Université Georgetown, de l'Université Howard, du NICHD et de RTI International.

Des interviews téléphoniques de base ont été réalisées en moyenne de trois à quatre semaines après le recrutement. Les femmes des deux groupes ne présentaient aucune différence importante au regard des caractéristiques de base (tableau 1). Les participantes pouvaient également être interviewées à trois autres occasions. Le suivi 1 (S1) a eu lieu entre 22 et 26 semaines d'AGE, le suivi 2 (S2), entre 34 et 38 semaines d'AGE, et l'interview post-partum (PP), de six à dix semaines après l'accouchement. Les taux de non-réponse variaient de 43 % pour le S1 à 20 % pour l'interview PP (tableau 2). Les données manquantes étaient en partie attribuables à l'inadmissibilité relativement à une interview en raison d'une inscription tardive, à un accouchement prématuré, etc. (figure 1). Le taux de données manquantes pour ce qui est des questions pour chaque interview se situait à moins de 2 %.

Caractéristiques	IISC (N=521)	SH (N=523)	Valeur de p (IISC c. SH)
Âge (en années)	24,4 ± 5,5	24,8 ± 5,3	0,25
AGE à l'interview de base (semaines)	19,3 ± 7,0	18,6 ± 6,8	0,11
Taille du ménage	3,9 ± 1,9	4,0 ± 1,9	0,24
Mariée/vit avec un conjoint	125 (24,0 %)	122 (23,3 %)	0,80
Diplôme d'études secondaires	362 (69,5 %)	366 (70,0 %)	0,86
Employée	185 (35,5 %)	196 (37,6 %)	0,49
Bénéficiaire de Medicaid	409 (79,1 %)	401 (76,8 %)	0,37
Consommation d'alcool pendant la grossesse (déclarée par l'intéressée)	111 (21,4 %)	112 (21,4 %)	0,98
Consommation de drogues illicites pendant la grossesse (déclarée par l'intéressée)	67 (12,9 %)	56 (10,7 %)	0,28
Tabagisme	106 (20,4 %)	92 (17,6 %)	0,26
EFTA	365 (71,4 %)	377 (73,4 %)	0,49
Dépression	229 (44,0 %)	234 (44,7 %)	0,80
VPI	169 (32,4 %)	167 (31,9 %)	0,86

Tableau 1 – Caractéristiques de base des participantes des groupes IISC et SH

Interview	Moment	% de participantes sans interview
Interview de base	3 à 4 semaines après le recrutement	0
Premier suivi (S1)	22 à 26 semaines d'AGE et au moins 4 semaines après l'interview de base	43
Second suivi (S2)	34 à 38 semaines d'AGE	31
Post-partum (PP)	6 à 10 semaines après l'accouchement	20

Tableau 2 – Pourcentage de participantes au projet DC-HOPE manquantes à chaque interview

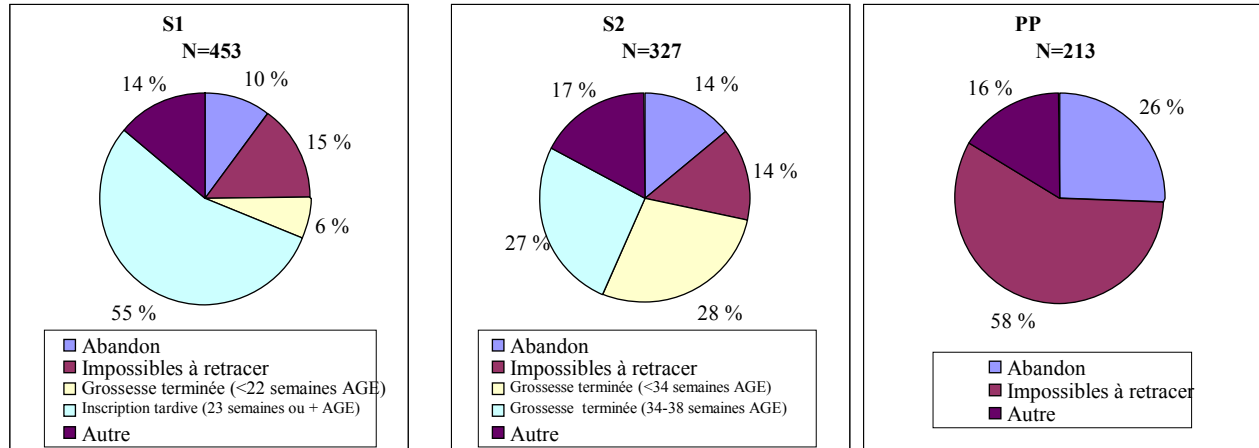


Figure 1 – Raisons pour lesquelles les interviews n’ont pas eu lieu

Les femmes qui ont manqué une des interviews de suivi pendant leur grossesse étaient proportionnellement plus nombreuses à être plus âgées (RC=1,04 pour +1 an, P=0,01), à avoir fumé plus de cent cigarettes dans leur vie (RC=1,53, P<0,01) et à être plus dépressive au moment de la première interview (RC=1,31 pour +1 unité dans le score moyen de la liste de contrôle des symptômes de Hopkins (Derogatis et coll., 1973; Derogatis et coll., 1974) (P=0,01). Les participantes qui ont manqué l’interview PP étaient plus susceptibles de vivre seules (RC=1,49, P=0,04), de recevoir une aide temporaire pour les familles nécessiteuses (RC=1,50, P<0,01) et d’avoir fumé plus de cent cigarettes dans leur vie (RC=1,41, P=0,03). De plus, les membres du groupe IISC qui n’avaient pas pris part aux interviews de suivi pendant leur grossesse ont assisté en moyenne à 3,5 moins de séances d’intervention que celles qui avaient pris part aux interviews (P<0,01). Les femmes du groupe IISC qui n’avaient pas participé à l’interview PP avaient assisté en moyenne à 3,1 fois moins de séances d’intervention que celles qui avaient participé à l’interview (P<0,01).

2. Méthodes

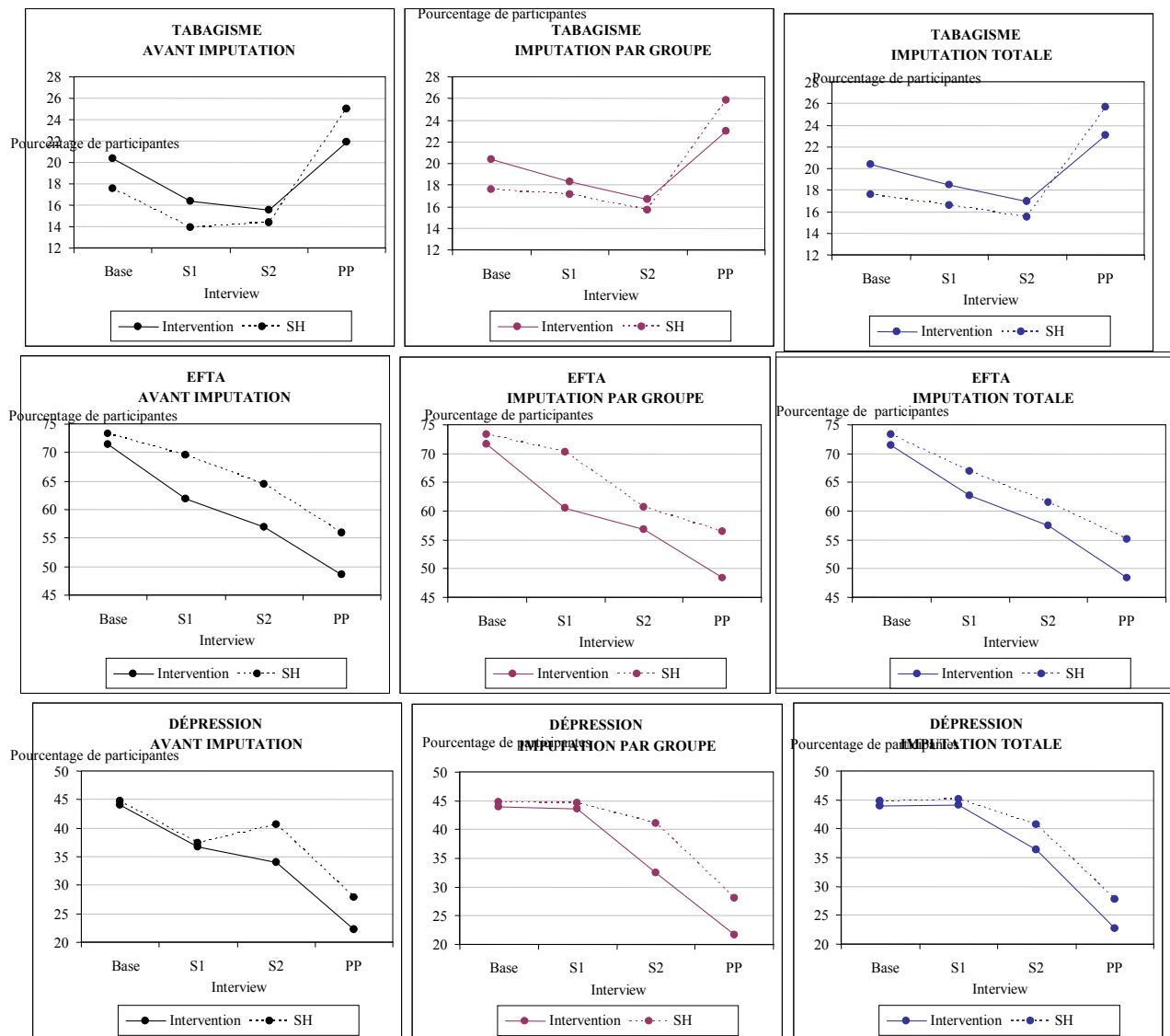
Concernant l’analyse des données, notre objectif était d’appliquer le principe de l’intention de traiter, dans le contexte duquel les participantes sont analysées selon le groupe de soins auquel elles ont été affectées aléatoirement, peu importe si elles ont véritablement fait l’objet d’une mesure d’intervention ou si elles se sont retirées de l’étude. Cette démarche permet d’éviter les biais pouvant être introduits lorsque des personnes non conformes sont exclues d’une analyse, et elle comporte une obligation de résultats pour tous les sujets. Pour que le principe de l’intention de traiter puisse être appliqué, la technique de l’imputation multiple a été retenue pour imputer les données manquantes des interviews. L’imputation multiple a été réalisée à l’aide du logiciel IVEware (logiciel d’imputation et d’estimation de la variance), mis au point par les responsables du programme des méthodes d’enquête, à l’institut pour la recherche sociale de l’Université du Michigan. Il est possible de télécharger le logiciel et la documentation connexe à <http://www.isr.umich.edu/src/smp/ive/>. Le logiciel permet d’effectuer l’imputation multiple en utilisant la méthode séquentielle d’imputation par régression décrite par Raghunathan et coll. (2001). Des modèles de régression linéaire, logistique, polytomique ou de Poisson sont utilisés pour imputer des valeurs manquantes continues, binaires, chiffrées ou nominales. Cinq ensembles complets de données étaient produits chaque fois qu’une imputation multiple était exécutée. La variabilité observée dans les ensembles de données imputées témoigne de l’incertitude inhérente aux prévisions de valeurs inconnues.

Le processus complet d’imputation multiple a été repris deux fois, dans deux optiques différentes. Il a tout d’abord été appliqué sur tout l’ensemble de données (imputation totale). Le modèle de régression qu’utilise le logiciel IVEware pour imputer des valeurs manquantes concernant une variable donnée englobe implicitement toutes les autres variables de l’ensemble de données comme prédicteurs. Toutefois, en raison de la multicollinéarité présente dans notre ensemble de données, nous n’avons retenu que les prédicteurs faisant augmenter de 0,135 la valeur de R au carré du modèle de régression. Dans la deuxième optique, les membres du groupe IISC ont été séparés de celles du groupe SH, et la technique d’imputation multiple a été appliquée individuellement à chaque groupe (imputation par groupe) à l’aide du même ensemble de prédicteurs. Cette méthode neutralisait implicitement l’effet du groupe de soins et les interactions

entre les groupes et les autres prédicteurs, ce qui permettait l'utilisation des modèles d'imputation plus complexes. Dans la méthode d'imputation par groupe, la valeur minimale de R au carré concernant les variables prédictives s'établissait à 0,155.

3. Résultats

En ce qui a trait à l'imputation par groupe (les données des deux groupes de soins ont été imputées séparément), les valeurs imputées avaient tendance à être comparables aux données non manquantes du groupe. Par contraste, dans le cas de l'imputation totale (les données des deux groupes de soins ont été imputées ensemble), les valeurs imputées pour le groupe IISC avaient tendance à ressembler davantage à celles du groupe SH (figure 2). Dans les deux cas, les taux de tabagisme étaient plus élevés qu'ils ne l'étaient dans les données initiales, ce qui est cohérent avec le fait que les femmes qui n'ont pas pris part aux interviews S1, S2 et PP étaient plus susceptibles d'avoir fumé plus de cent cigarettes dans leur vie. Dans le même ordre d'idées, les taux de dépression plus élevés dans les ensembles de données imputées témoignent du fait que les femmes qui étaient plus dépressives pendant l'interview de base étaient moins susceptibles de répondre aux questions des interviews de suivi.



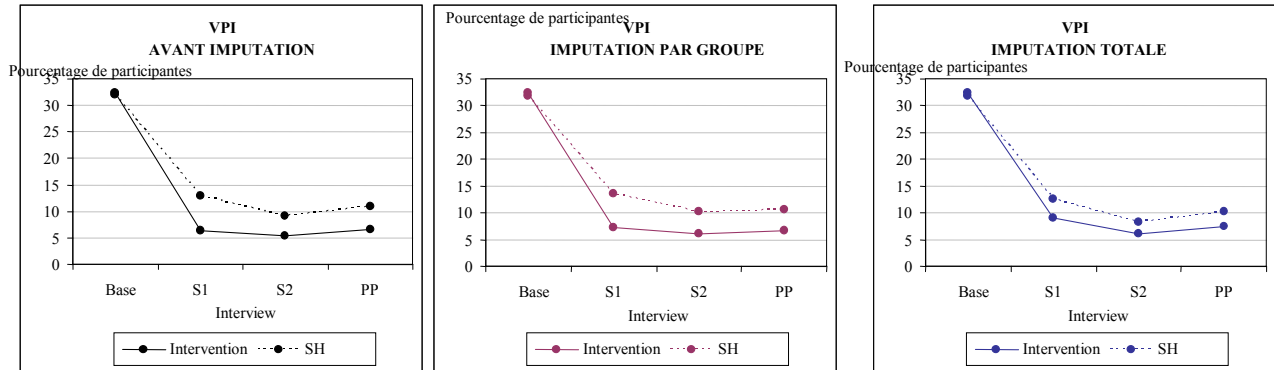


Figure 2 – Pourcentage de femmes déclarant des risques, ensembles de données avant et après imputation

Des modèles de régression ordinale ont été créés pour prévoir le nombre de risques déclarés par les participantes aux interviews S2 et PP. Le résultat comprenait quatre catégories : 0 risque, 1 risque, 2 risques et 3 ou 4 risques. Les caractéristiques de base énumérées au tableau 1 ont été intégrées à titre de covariables si les effets étaient significatifs dans le modèle de régression. L'analyse des données imputées multiples supposait l'exécution du modèle pour chacune des cinq imputations séparément et le regroupement ultérieur des résultats en une seule inférence globale qui neutralisait la variance entre les imputations et au sein de celles-ci (Rubin, 1987). La version 9.1 du logiciel SAS (SAS Institute, Inc., Cary, NC) a été utilisée pour toutes les analyses statistiques. La procédure MIANALYZE a d'autre part été utilisée pour obtenir l'inférence d'imputation multiple.

Qu'il s'agisse de l'imputation totale ou de l'imputation par groupe, il en résultait des effets d'intervention réduits par comparaison aux données avant imputation (voir le tableau 3). Pour ce qui est de l'imputation totale, la différence entre les groupes de soins était la plus petite.

Ensemble de données	Effet du groupe de soins (SH et IISC)			
	Nombre de risques à S2		Nombre de risques à PP	
	RC	P	RC	P
Avant imputation	1,65	<,01	1,90	<,01
Imputation par groupe	1,55	<,01	1,74	<,01
Imputation totale	1,36	0,04	1,60	<,01

Tableau 3 – Résultats du modèle de régression ordinale prédisant le fait de se trouver dans une catégorie à risque plus élevé (0, 1, 2 ou 3 et 4 risques)

On a eu recours à des méthodes longitudinales pour effectuer des tests relatifs à l'effet de l'intervention au fil du temps (S1-PP). Des modèles généralisés mixtes linéaires ont été créés pour chacun des quatre facteurs de risque, intégrant d'importantes covariables de l'interview de base. Pour les quatre facteurs de risque, le modèle de l'ensemble de données des groupes IISC et SH imputées ensemble (imputation totale) présentait le rapport de cotes (RC) le plus petit pour l'effet de l'intervention. Dans les modèles prédisant le tabagisme, le RC pour l'effet de l'intervention sur le tabagisme était le plus élevé dans les données avant imputation, mais l'effet était significatif uniquement dans les données imputées par groupe (voir le tableau 4). Et encore, l'effet n'était pas très significatif ($P=0,04$). L'effet de l'intervention sur l'EFTA était significatif dans les trois ensembles de données, quoique l'effet était le moins significatif lorsque les données des groupes de soins étaient imputées ensemble. Pour ce qui est de la dépression, seules les données imputées par groupe affichaient un effet de l'intervention significatif. L'effet de l'intervention sur la VPI était significatif dans les trois ensembles de données.

Ensemble de données	Effet du groupe de soins sur les risques au fil du temps (SH c. IISC)							
	Tabagisme		EFTA		Dépression		VPI	
	RC	P	RC	P	RC	P	RC	P
Avant imputation	1,56	0,75	1,44	<,01	1,37	0,03	1,73	<,01
Imputation par groupe	1,41	0,04	1,35	<,01	1,35	<,01	1,82	0,02
Imputation totale	1,24	0,18	1,22	0,04	1,22	0,07	1,48	<,01

Tableau 4 – Résultats des modèles généralisés mixtes linéaires prédisant les risques au fil du temps (S1-PP)

4. Conclusions

Les deux méthodes d'imputation des données ont produit des résultats quelque peu différents, mais des ressemblances étaient tout de même discernables. Tant la méthode de l'imputation totale que celle de l'imputation par groupe ont permis de conclure que les fumeuses et les participantes les plus déprimées étaient sous-représentées dans les interviews de suivi. Cette conclusion est encourageante, puisqu'elle indique que les covariables associées aux résultats manquants ont pu être repérées par le logiciel d'imputation et intégrées aux modèles de prévision utilisés pour estimer les valeurs manquantes.

Comme nous l'avons signalé plus haut, la méthode de l'imputation par groupe neutralise presque explicitement l'effet du groupe de soins et toutes les interactions possibles entre le groupe de soins et les autres covariables. L'inconvénient est que cette méthode suppose que les femmes du groupe IISC qui n'ont pas participé aux interviews de suivi étaient susceptibles d'avoir les mêmes comportements que celles qui ont répondu aux questions de suivi. Or, cette hypothèse peut fort bien être erronée. Bien qu'il soit possible d'ajuster les modèles d'imputation pour tenir compte des différences connues entre les répondantes et les non-répondantes, comme les différentes probabilités relativement au tabagisme ou à la dépression, il peut aussi y avoir d'autres différences non mesurées. L'établissement de modèles d'imputation distincts pour les femmes de chaque groupe de soins aurait pu se solder par une surestimation des effets de l'intervention. Ce biais a été potentiellement diminué par la présence d'une covariable concernant le nombre de séances d'intervention auxquelles les femmes ont assisté, mais cette variable n'a été prise en ligne de compte dans les modèles de régression que si elle faisait augmenter de 0,155 la valeur de R au carré. Il s'ensuit que la méthode de l'imputation par groupe représente le « scénario de réussite » pour l'effet de l'intervention qui pourrait être détecté s'il n'y avait pas eu de données manquantes.

Par contraste, la méthode de l'imputation totale pourrait se révéler trop prudente. Il faut se rappeler que les modèles de régression utilisés pour estimer les valeurs manquantes pour chaque variable tenaient uniquement compte des prédicteurs qui faisaient augmenter de 0,135 la valeur de R au carré. Donc, même si le groupe de soins et les interactions clés concernant le groupe de soins étaient des prédicteurs potentiels, s'ils ne respectaient pas le critère minimum relatif à la valeur de R au carré, ils n'étaient pas retenus dans les modèles de régression. Il est par conséquent possible que l'effet de l'intervention soit sous-estimé dans les données obtenues par imputation totale. D'autre part, la ressemblance des valeurs imputées pour les groupes IISC et SH pourrait indiquer que les non-répondantes dans le groupe IISC étaient moins susceptibles de profiter de l'avantage de l'intervention. Cette conclusion serait cohérente avec le fait que les femmes du groupe IISC qui ont assisté à un plus petit nombre de séances d'intervention étaient moins susceptibles de répondre aux questionnaires des interviews de suivi.

En résumé, la méthode de l'imputation totale donne probablement un ordre de grandeur plus prudent, mais aussi plus exact, de l'effet de l'intervention. Comme les femmes qui n'ont pas participé aux interviews de suivi étaient moins susceptibles d'avoir assisté aux séances d'intervention, tant les données imputées par groupe que les données avant imputation surestimaient probablement l'effet de faire partie du groupe IISC. La méthode de l'imputation totale est donc plus indiquée dans le contexte du principe de l'intention de traiter, qui évalue l'effet d'être assignée au hasard à un groupe de soins sans que soient pris en considération l'adhésion à ce groupe ou son retrait.

Références

- Derogatis, L. R., Lipman, R. S. et L. Coci (1973), "An outpatient psychiatric rating scale-preliminary report," *Psychopharmacology Bulletin*, pp. 13-28.
- Derogatis, L., Rickels, K., Uhlenhuth, E. H. et L. Covi (1974), "The Hopkins symptom checklist: a measure of primary symptom dimensions," in P. Pichot (ed.) *Psychological Measurements in Psychopharmacology: Problems in Psychopharmacology*, Basel: Karger, pp. 79-110.
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., et P. Solenberger (2001), "Une technique multidimensionnelle d'imputation multiple des valeurs manquantes à l'aide d'une séquence de modèles de régression", *Techniques d'enquête*, 27, pp. 91-103.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.
- Thornberry, J. S., El-Khorazaty, M. N., El-Mohandes, A. E., Johnson, A. A., Joseph, J. G., Sivasubramanian, K. N., Kiely, M., et the NIH-DC Initiative to Reduce Infant Mortality in Minority Populations (2003), "Acceptability of Audio Computer-Assisted Self-Interview (A-CASI) as a screening tool for risk factors in women receiving prenatal care," article présenté au Pediatric Academic Societies' Annual Meeting, Seattle, Washington.