

Techniques d'enquête

Approches fréquentiste et bayésienne pour comparer les composantes de l'écart intervieweurs dans deux groupes d'intervieweurs d'enquête

par Brady T. West et Michael R. Elliott

Date de diffusion : 19 décembre 2014



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-877-287-4369 |

Programme des services de dépôt

- | | |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur | 1-800-565-7757 |

Comment accéder à ce produit

Le produit no 12-001-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Publication autorisée par le ministre responsable de
Statistique Canada

© Ministre de l'Industrie, 2014

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada (<http://www.statcan.gc.ca/reference/copyright-droit-auteur-fra.htm>).

This publication is also available in English.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, ses entreprises, ses administrations et les autres établissements. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- P provisoire
- r révisé
- X confidentiel en vertu des dispositions de la *Loi sur la statistique*
- E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$)

Approches fréquentiste et bayésienne pour comparer les composantes de l'écart intervieweurs dans deux groupes d'intervieweurs d'enquête

Brady T. West et Michael R. Elliott¹

Résumé

Les méthodologistes d'enquête étudient depuis longtemps les effets des intervieweurs sur la variance des estimations d'enquête. Les modèles statistiques tenant compte des effets aléatoires des intervieweurs sont souvent intégrés à ce genre d'études, et l'intérêt de la recherche repose sur l'ampleur de la composante de la variance de l'intervieweur. Une question peut se poser au cours d'une étude méthodologique : différents groupes d'intervieweurs (p. ex. ceux ayant de l'expérience relative à une enquête donnée par rapport aux nouvelles recrues, ou les intervieweurs IPAO par rapport aux intervieweurs ITAO) ont-ils des composantes de variance considérablement différentes dans ces modèles? Des écarts importants peuvent indiquer un besoin de formation supplémentaire pour certains sous-groupes, les propriétés moins optimales de différents modes ou styles d'interview pour certaines questions d'enquête (en ce qui concerne l'erreur quadratique moyenne globale des estimations d'enquête). Les chercheurs d'enquête désirant des réponses à ces types de questions disposent de différents outils statistiques. Le présent article cherche à fournir un aperçu des approches fréquentiste et bayésienne de rechange de la comparaison des composantes de la variance dans différents groupes d'intervieweurs d'enquête, au moyen d'un cadre de modélisation linéaire généralisée hiérarchique qui tient compte de différents types de variables d'enquête. Nous considérons d'abord les avantages et les limites de chaque approche, en comparant les méthodes utilisées pour l'estimation et l'inférence. Nous présentons ensuite une étude de simulation, en évaluant de façon empirique la capacité de chaque approche d'estimer efficacement les différences entre les composantes de la variance. Nous appliquons alors les deux approches à une analyse des données d'enquête réelles recueillies dans le cadre de la *National Survey of Family Growth* (NSFG) aux États-Unis. Nous concluons que les deux approches ont tendance à donner des inférences très semblables et nous présentons des suggestions à mettre en pratique, compte tenu des différences subtiles observées.

Mots clés : Écart intervieweurs; analyse bayésienne; modèles linéaires généralisés hiérarchiques; test du rapport des vraisemblances.

1 Introduction

La variance entre intervieweurs des méthodes d'enquête (p. ex. West, Kreuter et Jaenichen 2013; West et Olson 2010; Gabler et Lahiri 2009; O'Muircheartaigh et Campanelli 1998; Biemer et Trewin 1997; Kish 1962) se produit lorsque les réponses à une enquête obtenues par un intervieweur donné se ressemblent davantage que les réponses recueillies par différents intervieweurs. La variance entre intervieweurs peut augmenter la variance des estimations d'enquête des moyennes, et peut se produire à cause d'écarts de réponse corrélés introduits par un intervieweur (p. ex. Biemer et Trewin 1997), compte tenu de la complexité des questions de l'enquête (p. ex. Collins et Butcher 1982) ou des interactions entre l'intervieweur et le répondant (p. ex. Mangione, Fowler et Louis 1992), ou de la variance de l'erreur de non-réponse entre les intervieweurs (West et coll. 2013; Lynn, Kaminska et Goldstein 2011; West et Olson 2010).

Les organismes de recherche par sondages forment les intervieweurs de manière à éliminer cette composante de la variance des estimations d'enquête, qui est parfois plus grande que la composante de la

1. Brady T. West, Survey Methodology Program, Institute for Social Research, 426, rue Thompson, Ann Arbor, MI, 48106, courriel : bwest@umich.edu; Michael R. Elliott, Survey Methodology Program, Institute for Social Research, 426, rue Thompson, Ann Arbor, MI, 48106, courriel : mrelliot@umich.edu.

variance due à l'échantillonnage en grappes (Schnell et Kreuter 2005). En fait, une composante de l'écart intervieweurs ne peut jamais être égale à 0 (ce qui supposerait que les moyennes de la variable d'intérêt sont identiques pour tous les intervieweurs), mais les gestionnaires d'enquête cherchent à réduire cette composante au moyen de formation spécialisée des intervieweurs. Par exemple, les intervieweurs peuvent s'exercer à poser certaines questions sous la supervision directe du personnel de formation, avant de recevoir des commentaires sur toute variance de l'administration observée par le personnel (afin de normaliser l'administration; voir Fowler et Mangione 1990). Dans certains plans non imbriqués, où les intervieweurs sont généralement affectés exclusivement à une seule zone d'échantillonnage primaire (p. ex. la National Survey of Family Growth aux États-Unis; voir Lepkowski, Mosher, Davis, Groves et Van Hoewyk 2010), les effets de l'intervieweur et les effets de la zone sont confondus, ce qui empêche l'estimation de la variance des estimations d'enquête qui est attribuable exclusivement aux intervieweurs. Les plans d'échantillonnage imbriqués astucieux (Mahalanobis 1946) permettent aux intervieweurs de travailler dans plusieurs zones d'échantillonnage, et dans ces cas, des modèles multiniveaux recoupés peuvent être utilisés pour estimer les composantes de la variance due aux intervieweurs et aux zones (p. ex. Durrant, Groves, Staetsky et Steele 2010; Gabler et Lahiri 2009; Schnell et Kreuter 2005; O'Muircheartaigh et Campanelli 1999; O'Muircheartaigh et Campanelli 1998).

En général, l'estimation de l'ampleur globale de l'écart intervieweurs en ce qui concerne les mesures d'une variable d'enquête donnée ou le résultat du processus de collecte des données est un exercice utile pour les praticiens d'enquête. Si des sous-échantillons aléatoires d'unités d'échantillonnage sont attribués aux intervieweurs après une conception imbriquée, on peut estimer la composante de la variance due aux intervieweurs et par la suite, les effets particuliers des intervieweurs sur la variance d'une moyenne d'enquête estimée (p. ex. Groves 2004, p. 364). Les grosses estimations peuvent indiquer des difficultés de mesure potentielles éprouvées par certains intervieweurs, ou une réussite différentielle possible du recrutement de types particuliers d'unités échantillonnées. Compte tenu d'une estimation relativement importante d'une composante de l'écart intervieweurs et d'un test statistique approprié indiquant que la composante est considérablement plus grande que zéro (ou « non négligeable », étant donné que les composantes de la variance ne peuvent pas en principe être exactement égales à zéro; voir Zhang et Lin 2010), les gestionnaires d'enquête peuvent utiliser diverses méthodes pour calculer les prévisions des effets aléatoires associés aux intervieweurs individuels, et pour cerner les intervieweurs qui éprouvent peut-être des difficultés avec certains aspects du processus de collecte des données.

Bien que l'estimation des composantes de l'écart intervieweurs et les rajustements subséquents de la formation des intervieweurs et des protocoles de collecte des données font partie de la documentation des méthodes d'enquête depuis longtemps (voir Schaeffer, Dykema et Maynard 2010, pour un examen récent), aucune étude des méthodes d'enquête à ce jour n'a examiné les approches de rechange qui sont offertes aux chercheurs d'enquête pour *comparer* les composantes de la variance dans deux groupes indépendants d'intervieweurs d'enquête. En général, les approches statistiques de rechange sont disponibles pour estimer les composantes de l'écart intervieweurs, et les estimations (et les inférences correspondantes au sujet des composantes de la variance) peuvent être sensibles à la méthodologie d'estimation qu'utilise un chercheur d'enquête. Il en va de même pour les chercheurs d'enquête qui peuvent souhaiter comparer les composantes de la variance associées à différents groupes d'intervieweurs, pour différentes raisons (p. ex. cerner les groupes qui ont besoin de plus de formation ou trouver des modes plus efficaces pour certains types de questions) : il existe différentes approches statistiques pour

effectuer ces types de comparaisons, et les inférences au sujet des différences peuvent varier en fonction de l'approche utilisée. Dans le présent document, nous cherchons à évaluer des approches fréquentistes et bayésiennes de rechange afin de faire des inférences à propos des fluctuations des composantes de la variance entre deux groupes indépendants d'intervieweurs d'enquête, et à fournir des conseils pratiques aux chercheurs d'enquête qui s'intéressent à ce type d'analyse.

Le document est structuré comme suit. À la section 2, nous présentons le cadre de modélisation général qui permet de faire ces comparaisons des composantes de l'écart intervieweurs pour les variables d'enquête normales et non normales (p. ex. binaire, compte), et nous passons en revue la documentation existante en comparant les approches fréquentistes et bayésiennes de l'estimation et de l'inférence, en décrivant les avantages et inconvénients de ces approches. Nous présentons ensuite une étude de simulation à la section 3, en évaluant la capacité des deux approches d'estimer efficacement les différences entre les composantes de la variance des deux groupes hypothétiques d'intervieweurs. La section 4 applique les deux approches aux données d'enquête réelles recueillies dans le cadre de la National Survey of Family Growth (NSFG) des États-Unis (Lepkowski et coll. 2010; Groves, Mosher, Lepkowski et Kirgis 2009). Enfin, la section 5 présente le mot de la fin, des suggestions aux praticiens et des orientations pour les recherches à venir. Nous incluons à l'annexe le code SAS, R et WinBUGS, que les lecteurs peuvent utiliser pour mettre en œuvre les deux approches.

2 Approches de rechange pour comparer les composantes de la variance dans les modèles linéaires généralisés hiérarchiques

Nous examinons d'abord une catégorie générale de modèles que les chercheurs d'enquête peuvent utiliser pour comparer les composantes de la variance dans différents groupes d'intervieweurs. Les modèles linéaires généralisés hiérarchiques (MLGH) sont des outils analytiques flexibles qui peuvent être utilisés pour modéliser des observations pour les variables d'enquête d'intérêt normales et non normales (p. ex. binaire, compte), où les observations recueillies par un même intervieweur ne peuvent pas être considérées comme indépendantes (Raudenbush et Bryk 2002; Goldstein 1995). Nous examinons des approches de rechange pour faire des inférences au sujet des composantes de l'écart intervieweurs dans une catégorie particulière de MLGH, où les composantes de l'écart intervieweurs pour les deux groupes indépendants d'intervieweurs définis par une caractéristique connue de l'intervieweur n'ont pas à être égales. Ce type de MLGH peut être écrit comme suit

$$\begin{aligned} g\left(E\left[y_{ij} \mid u_i\right]\right) &= \beta_0 + \beta_1 I(\text{Groupe} = 1)_i + u_{i(1)} I(\text{Groupe} = 1)_i + u_{i(2)} I(\text{Groupe} = 2)_i \\ u_{i(1)} &\sim N\left(0, \tau_1^2\right), u_{i(2)} \sim N\left(0, \tau_2^2\right), \end{aligned} \quad (2.1)$$

où $g(x)$ est la fonction de lien appariant une transformation de la valeur prévue de la variable dépendante, y_{ij} , à la combinaison linéaire des effets fixes et aléatoires (p. ex. $g(x) = \log\left[\frac{x}{1-x}\right]$ pour une répartition Bernoulli présumée [résultat binaire], $g(x) = \log(x)$ pour une répartition de Poisson présumée [résultat du compte]), i est un indice pour l'intervieweur, j est un indice pour le répondant associé à un seul intervieweur, et $I(\bullet)$ représente une variable indicatrice, équivalant à 1 si la

condition à l'intérieur des parenthèses est vraie et 0 autrement. Les effets aléatoires des intervieweurs du groupe 1, $u_{i(1)}$, sont présumés suivre une répartition normale, avec la moyenne 0 et la variance τ_1^2 , tandis que les effets aléatoires des intervieweurs du groupe 2, $u_{i(2)}$, sont présumés suivre une répartition normale avec la moyenne 0 et la variance τ_2^2 . D'autres répartitions peuvent être postulées pour les effets aléatoires, et le modèle général dans (2.1) peut tenir compte de la surdispersion de la variable dépendante observée par rapport à la répartition postulée pour cette variable. L'aspect clé de la spécification dans (2.1) est que les effets aléatoires pour différents groupes d'intervieweurs ont des variances *différentes*. Le paramètre des effets fixes β_1 dans (2.1) représente un effet fixe du groupe 1 sur le résultat par rapport au groupe 2 dans le MLGH, et des effets fixes d'autres covariables peuvent être inclus facilement. De même, des sous-groupes supplémentaires d'intervieweurs peuvent être considérés au moyen d'effets aléatoires supplémentaires $u_{i(k)}$, pour $k > 2$. L'intérêt analytique repose sur l'ampleur de la différence entre les composantes de la variance.

Les modèles de la forme dans (2.1) peuvent être appliqués lorsque les études méthodologiques sont conçues de manière à comparer deux groupes différents d'intervieweurs en ce qui concerne leurs composantes de la variance. Par exemple, il y a un débat dans la documentation sur les méthodes d'enquête en vue de déterminer si les intervieweurs devraient utiliser des interviews normalisées ou interactives. Les partisans de l'interview normalisée soutiennent que tous les intervieweurs devraient administrer les enquêtes exactement de la même façon, afin de permettre aux répondants de décoder les questions comme ils l'entendent (p. ex. Fowler et Mangione 1990). D'autres recherches ont démontré qu'une méthode d'interview plus flexible axée sur un style interactif pourrait accroître la compréhension des répondants des questions d'enquête et réduire l'erreur de mesure (p. ex. Schober et Conrad 1997). Pour tester l'hypothèse voulant qu'un style d'interview entraîne une plus faible variance entre intervieweurs, un chercheur peut randomiser les intervieweurs en deux groupes ayant reçu une formation dans les deux styles différents, recueillir des données d'enquête sur diverses variables, puis intégrer le modèle (2.1), y compris les variables indicatrices pour les deux groupes d'intervieweurs. Cette même approche pourrait être utilisée pour comparer les composantes de l'écart intervieweurs en deux groupes d'intervieweurs affectés aléatoirement à différents modes de collecte des données (p. ex. IPAO ou ITAO). À ce jour, aucune étude publiée n'a tenté d'effectuer ce genre de comparaisons, qui sont toutefois importantes pour comprendre les effets généraux de ces décisions relatives au plan sur l'erreur quadratique moyenne (EQM) des estimations d'enquête.

Les approches fréquentistes de l'estimation des paramètres dans les MLGH reposent sur diverses approches numériques ou théoriques de l'approximation de fonctions complexes de probabilité, en particulier pour les modèles comme (2.1), qui supposent des structures complexes des effets aléatoires (p. ex. Faraway 2006, chapitre 10; Molenberghs et Verbeke 2005). En général, les inférences sont basées sur ces approches approximatives axées sur la probabilité, qui comprennent la pseudo-vraisemblance résiduelle (qui diffère de l'approche de l'estimation du pseudo-maximum de vraisemblance élaborée par Binder (1983) pour les analyses fondées sur le plan des données d'enquêtes-échantillons complexes), la quasi-vraisemblance pénalisée et le maximum de vraisemblance basé sur une approximation de Laplace. Des travaux antérieurs ont donné des résultats de simulation favorables pour l'approche de pseudo-vraisemblance résiduelle, qui indiquent une estimation presque sans biais des composantes de la variance dans un MLGH comparativement au maximum de vraisemblance au moyen de l'approximation de

Laplace ou de la quadrature adaptative (Pinheiro et Chao 2006). Ces résultats ressemblent au cas de l'estimation du maximum de vraisemblance restreint (REML, pour *restricted maximum likelihood*) dans un modèle pour une variable de résultat normalement réparti. Pour les variables de résultat binaire, les techniques de quasi-vraisemblance marginale ou pénalisée peuvent donner lieu à un biais vers le bas dans les estimations des paramètres et à des problèmes de convergence, et les approches exclusivement bayésiennes peuvent avoir des propriétés favorables dans ce cas (Browne et Draper 2006; Rodriguez et Goldman 2001). Nous considérons donc l'approche de pseudo-vraisemblance résiduelle dans les simulations et les applications présentées dans cette étude, et comparons cette approche à une approche exclusivement bayésienne.

Il existe deux approches pour faire une inférence à propos des différences entre les composantes de la variance dans le contexte fréquentiste. La première approche consiste à tester l'hypothèse nulle voulant que $\tau_1^2 = \tau_2^2$, par rapport à l'hypothèse de rechange voulant que $\tau_1^2 \neq \tau_2^2$. En principe, il s'agit d'un simple test d'hypothèse à effectuer au moyen de la méthodologie fréquentiste, puisque l'hypothèse nulle définit une contrainte d'égalité au lieu d'établir un paramètre pour une valeur à la limite d'un espace des paramètres. Le modèle sous l'hypothèse nulle est imbriqué dans le modèle sous l'hypothèse de rechange, où $\tau_2^2 = \tau_1^2 + k$. L'hypothèse nulle peut donc être réécrite $k = 0$, au lieu de $k \neq 0$. Un essai statistique est calculé par l'intégration d'une version limitée du modèle dans (2.1), les composantes de la variance de l'effet aléatoire dans les deux groupes étant réputées égales, puis par l'intégration du modèle à la forme plus générale dans (2.1). La différence positive entre les valeurs approximatives du logarithme du rapport de vraisemblance -2 de ces deux modèles est alors calculée, et appelée répartition du khi carré avec un degré de liberté.

La deuxième méthode nécessite le calcul de la différence entre les estimations pseudo-MV, $\hat{\tau}_1 - \hat{\tau}_2$, et un intervalle de confiance connexe de type Wald de 95 % pour la différence, donnée par $\hat{\tau}_1 - \hat{\tau}_2 \pm 1,96\sqrt{\hat{\text{var}}(\hat{\tau}_1) + \hat{\text{var}}(\hat{\tau}_2) - 2\hat{\text{cov}}(\hat{\tau}_1, \hat{\tau}_2)}$. Cet intervalle nécessite des estimations asymptotiques des variances et des covariances des deux composantes estimées de la variance, qui sont calculées en fonction de la matrice d'Hessien (dérivée seconde) de la fonction objective utilisée pour la procédure d'estimation du maximum de vraisemblance. Si l'intervalle de Wald obtenu comprend zéro, on pourrait conclure que les données ne permettent pas de rejeter l'hypothèse nulle. Les intervalles de confiance pour les différences entre les composantes de la variance peuvent également être calculés au moyen des inversions des tests de vraisemblance des profils (p. ex. Viechtbauer 2007), mais le logiciel standard ne comprend pas d'options pour mettre en œuvre cette procédure (à notre connaissance).

Ces deux approches fréquentistes pour faire une inférence au sujet des différences entre les composantes de l'écart intervieweurs ont des limites. Lorsque le nombre d'intervieweurs dans chaque groupe est petit (disons moins de 30; voir Hox 1998, pour une discussion), les résultats asymptotiques pour le test du rapport des vraisemblances (Zhang et Lin 2010) risquent de ne plus tenir. Les méthodes fréquentistes (maximum de vraisemblance) ont également tendance à surévaluer la précision des estimations, puisqu'elles ne tiennent pas compte de l'incertitude des estimations des composantes de la variance (Carlin et Louis 2009, p. 335-336), qui est particulièrement problématique pour les petits échantillons (Goldstein 1995, p. 23). Les approches bayésiennes permettent aux analystes de disposer les répartitions a priori des composantes de la variance pour tenir compte de cette incertitude, contrairement aux approches fréquentistes. En outre, Molenberghs et Verbeke (2005, p. 277) soutiennent que les tests du rapport des vraisemblances ne devraient pas être utilisés pour tester des hypothèses lorsque les modèles

sont intégrés au moyen de méthodes de pseudo-vraisemblance. Les méthodes d'estimation approximative du maximum de vraisemblance peuvent également donner lieu à des estimations invalides (c.-à-d. négatives) des composantes de la variance dans ces modèles lorsque les composantes de la variance sont très petites. Les logiciels qui n'utilisent pas de procédures d'estimation limitant ces composantes de la variance à des valeurs supérieures à zéro répondent généralement à ce problème en établissant des estimations négatives des composantes de la variance égales à zéro (sans l'erreur type qui les accompagne), ce qui empêche le calcul de l'intervalle de confiance de type Wald susmentionné.

Une approche bayésienne pour intégrer les MLGH décrits dans (2.1) utilise l'échantillonneur de Gibbs basé sur les MCMC et la méthodologie d'échantillonnage-rejet d'adaptation (Gilks et Wild 1992) pour simuler des tirages de la répartition a posteriori pour les paramètres du modèle défini dans (2.1). En général, les répartitions a posteriori pour les paramètres dans un MLGH n'ont pas de formes de répartition connues et doivent être simulées (Gelman, Carlin, Stern et Rubin 2004, section 16.4). Les valeurs a priori diffuses et non informatives pour les effets fixes et les composantes de la variance dans (2.1) peuvent être précisées pour les simulations, afin de laisser les données fournir le plus d'information au sujet des répartitions a posteriori des paramètres (Gelman et Hill 2007; Gelman 2006, section 7). Cette approche donne lieu à des inférences basées sur les tirages simulés des répartitions a posteriori marginales des deux paramètres des effets fixes, les deux paramètres de la variance, les effets aléatoires des intervieweurs et toute fonction de ces paramètres. Cette étude s'intéresse à la répartition a posteriori marginale de la différence entre les variances des effets aléatoires de deux groupes d'intervieweurs définis par une caractéristique connue au niveau de l'intervieweur, calculée au moyen des tirages simulés des deux composantes de la variance.

Étant donné que les vérifications d'hypothèses traditionnelles ne sont pas utiles dans le contexte bayésien, l'inférence bayésienne sera axée sur la différence entre les composantes de l'écart intervieweurs. L'inférence pour la différence est basée sur plusieurs milliers de tirages pour les deux composantes de la variance de la répartition a posteriori combinée estimée au moyen de la formule d'échantillonnage de Gibbs. Pour chaque tirage d des deux composantes de la variance, la différence entre les composantes de la variance, définie comme $\tau_1^{2(d)} - \tau_2^{2(d)}$, peut être calculée. Les inférences seront donc basées sur la répartition marginale de ces différences, sans tenir compte des effets aléatoires des intervieweurs et des autres paramètres dérangement. La médiane et les quantiles 0,025 et 0,975 (pour un ensemble crédible de 95 %) des différences simulées entre les deux composantes de la variance seront calculés en fonction du nombre réel de tirages de simulation des deux composantes de la variance à partir de la répartition a posteriori combinée estimée. Dans une analyse donnée, plusieurs milliers de tirages de la répartition a posteriori peuvent être générés au moyen de la formule d'échantillonnage de Gibbs, un grand nombre de tirages initiaux étant éliminés en tant que tirages de rodage, et le nombre réel de tirages de simulation sera calculé en fonction du nombre de tirages de rodage (Gelman et Hill 2007, chapitre 16). Si l'ensemble crédible de 95 % obtenu comprend 0, les données supporteront l'hypothèse des deux groupes ayant des composantes égales de la variance. Si l'ensemble crédible de 95 % ne comprend pas 0, les données supporteront l'hypothèse des deux groupes ayant des variances différentes, une médiane positive suggérant que le groupe 1 a la composante de variance plus élevée. L'inférence pour les deux effets fixes peut suivre une approche semblable.

Le fait de se concentrer sur les tirages des deux composantes de la variance à partir de la répartition a posteriori combinée complète (et leurs différences) et de faire fi des tirages des effets aléatoires des

intervieweurs et des effets fixes a pour conséquence d'intégrer ces autres paramètres de la répartition a posteriori combinée. Cette approche bayésienne offre donc une méthode pratique pour simuler les tirages à partir de la répartition marginale d'un paramètre complexe (la différence entre les deux composantes de la variance) et pour calculer un ensemble crédible de 95 % pour ce paramètre. Ces estimations peuvent également être obtenues dans l'approche fréquentiste, comme susmentionné, mais l'approche bayésienne n'exige pas de suppositions asymptotiques et comprend la variabilité des composantes de la variance estimée dans le calcul des ensembles crédibles de 95 % au moyen des tirages simulés.

Plusieurs (habituellement trois) chaînes de Markov peuvent être exécutées en parallèle selon l'algorithme d'échantillonnage itératif de Gibbs pour simuler des trajets aléatoires dans l'espace de la répartition a posteriori combinée. La statistique \hat{R} de Gelman-Rubin, qui représente (à peu près) la racine carrée de la variance de l'amalgame des chaînes divisée par la variance moyenne dans la chaîne (Gelman et Rubin 1992), peut être utilisée pour évaluer la convergence (ou la combinaison) des chaînes pour chaque paramètre. Les valeurs inférieures à 1,1 dans cette statistique peuvent être considérées comme un indice de convergence des chaînes pour un paramètre donné. Les tirages a posteriori des paramètres peuvent être dérivés des trois chaînes combinées pour générer la taille finale de l'échantillon réel de tirages utilisés pour les inférences.

L'approche bayésienne susmentionnée a par ailleurs ses limites. La sélection des répartitions a priori utilisées pour calculer la répartition a posteriori pour les paramètres dans (2.1) est essentiellement arbitraire et dépend des choix d'un analyste donné et de la quantité d'information a priori disponible. De plus, le choix de la répartition a priori peut devenir essentiel lorsqu'il y a un petit nombre d'intervieweurs (disons moins de 20), où différentes distributions a priori peuvent entraîner des inférences très différentes au sujet des composantes de la variance (Lambert, Sutton, Burton, Abrams et Jones 2005); l'utilisation de l'information a priori au sujet des composantes de la variance peut accroître l'efficacité relative à l'utilisation de distributions a priori non informatives dans ces cas. L'erreur de spécification du modèle est également une nette possibilité, selon la variable de l'enquête modélisée, qui s'avère aussi une faiblesse de l'approche fréquentiste. La demande relative aux calculs peut également être problématique dans l'approche bayésienne (méthode d'échantillonnage de Gibbs) (Browne et Draper 2006), en particulier lorsque l'on souhaite faire des comparaisons des composantes de l'écart intervieweurs pour un grand nombre de variables d'enquête (avec des répartitions potentiellement différentes) et qu'il y a un nombre relativement élevé d'intervieweurs; ce n'est pas nécessairement aussi problématique compte tenu des progrès réalisés récemment en matière de rapidité du matériel et de l'efficacité des algorithmes. Enfin, les analystes ne sont peut-être pas à l'aise avec les logiciels disponibles pour les approches bayésiennes, ce qui fait qu'il pourrait y avoir une courbe d'apprentissage associée à la mise en œuvre de cette approche.

Plusieurs articles antérieurs ont comparé ces approches fréquentistes et bayésiennes de rechange au moyen d'études de simulation. Chaloner (1987) a envisagé des modèles d'analyse de la variance à un facteur avec effets aléatoires pour les données déséquilibrées (à l'instar du cas dans cette étude, où les intervieweurs ont des charges de travail variables), et observé des valeurs empiriques inférieures de l'EQM pour les modes a posteriori des composantes de la variance selon l'approche bayésienne et au moyen des distributions a priori non informatives par rapport à l'approche fréquentiste (maximum de vraisemblance). Van Tassel et Van Vleck (1996) ont déclaré que la formule d'échantillonnage de Gibbs (au moyen des répartitions préalables informatives ou non informatives) et le REML produisent tous deux des estimations empiriques non biaisées des composantes de la variance qui ont tendance à être très

semblables. Browne et Draper (2006) ont également constaté que les deux approches peuvent donner lieu à des estimations non biaisées, la nature plus « automatique » des approches fréquentistes étant un atout intéressant. Dans le contexte de la prédiction des moyennes pour les petites régions au moyen de modèles avec effets aléatoires des régions, Singh, Stukel et Pfeffermann (1998) ont déclaré que les approximations bayésiennes de l'EQM pour les prédictions ont de bonnes propriétés fréquentistes, mais que la méthode bayésienne a tendance à produire des biais fréquentistes et des EQM de prédiction plus marquées que les méthodes fréquentistes. Farrell (2000) a constaté que l'approche bayésienne entraînait des prédictions légèrement plus précises des proportions des petites régions, comportant peu de variations des taux de couverture entre les deux approches. Ugarte, Goicoa et Militino (2009) ont également conclu que les deux approches donnaient des résultats très similaires dans une application nécessitant la détection des zones à haut risque de maladie. Ces auteurs soulignent que la simplicité relative du calcul de l'approche fréquentiste est intéressante compte tenu de ces résultats. En général, d'après la documentation dans ce domaine, nous prévoyons un rendement similaire des deux méthodes dans le cas d'une comparaison des composantes de l'écart intervieweurs, et nous évaluons cette attente au moyen d'une étude de simulation (section 3).

Bien qu'il y ait de nombreuses procédures logicielles pour l'intégration de modèles multiniveaux et de composantes de la variance d'estimation au moyen des méthodes fréquentistes et bayésiennes (voir West et Galecki 2011 pour un examen), l'approche fréquentiste de la comparaison particulière des composantes de la variance dont il est question dans le présent document n'est facilement mise en œuvre que dans la procédure GLIMMIX de SAS/STAT (SAS 2010), au moyen de l'énoncé COVTEST avec l'option HOMOGENEITY (qui présume qu'une variable GROUP a été précisée dans l'énoncé RANDOM, indiquant différents groupes de grappes ayant des effets aléatoires découlant des différentes répartitions). Au moment de la rédaction, nous ne connaissons pas d'autres procédures permettant la mise en œuvre facile de l'approche comparative fréquentiste. Un exemple de code pouvant être utilisé pour l'intégration de ces modèles au moyen de la procédure GLIMMIX est disponible à l'annexe. L'approche bayésienne pour comparer les composantes de la variance peut être mise en œuvre dans le logiciel BUGS (inférence bayésienne au moyen de la méthode d'échantillonnage de Gibbs) (voir la bibliographie pour obtenir plus de détails). Nous incluons également dans l'annexe un exemple de code qui met en œuvre cette approche en ouvrant WinBUGS à partir de R dans l'annexe.

3 Étude de simulation

Nous avons réalisé une petite étude de simulation en vue d'examiner les propriétés empiriques de ces deux approches de rechange. Les données des deux variables d'enquête hypothétiques d'intérêt (répartition normale et répartition de Bernoulli) ont été simulées selon les deux modèles de super-population suivants :

$$\begin{aligned}
 y_{ij} &= 45 + 5 \times I(\text{Groupe} = 2)_i + u_{i(1)} I(\text{Groupe} = 1)_i + u_{i(2)} I(\text{Groupe} = 2)_i + \varepsilon_{ij} \\
 u_{i(1)} &\sim N(0,1), \quad u_{i(2)} \sim N(0,2), \quad \varepsilon_{ij} \sim N(0,64)
 \end{aligned}
 \tag{3.1}$$

$$P(y_{ij} = 1) = \frac{\exp[-1 + u_{i(1)}I(\text{Groupe} = 1)_i + u_{i(2)}I(\text{Groupe} = 2)_i]}{1 + \exp[-1 + u_{i(1)}I(\text{Groupe} = 1)_i + u_{i(2)}I(\text{Groupe} = 2)_i]} \quad (3.2)$$

$$u_{i(1)} \sim N(0; 0,03), u_{i(2)} \sim N(0; 0,13).$$

La notation utilisée ici est conforme à celle qui a été utilisée à (2.1). Les valeurs de la deuxième variable de Bernoulli ont été générées pour les cas hypothétiques selon le modèle de régression logistique indiqué à (3.2). Pour obtenir la variable de Bernoulli observée, un tirage aléatoire a été obtenu à partir d'une répartition UNIFORME(0,1), et la variable a été établie à 1 si le tirage aléatoire était inférieur ou égal à la probabilité prédite, et à 0 autrement. Pour un groupe hypothétique d'intervieweurs à la fois, les effets aléatoires des intervieweurs ont été tirés, et les valeurs pour les cas de chaque intervieweur ont été générées en fonction du modèle précisé.

Nous avons généré 200 échantillons de cas hypothétiques et de données simulées pour chaque variable, 50 intervieweurs hypothétiques dans un groupe recueillant des données à partir de 50 cas hypothétiques chacun ($n = 2\,500$ pour chaque groupe d'intervieweurs). Nous avons ensuite généré 200 autres échantillons dans un scénario de petit échantillon, 20 intervieweurs dans chaque groupe recueillant des données à partir de 10 cas hypothétiques chacun ($n = 200$ pour chaque groupe d'intervieweurs). Les choix des composantes de la variance à (3.1) correspondent aux corrélations intra-intervieweur de 0,015 et 0,030 pour les deux groupes hypothétiques d'intervieweurs, tandis que les choix des composantes de la variance à (3.2) correspondent aux corrélations intra-intervieweur de 0,009 et 0,038. Toutes ces valeurs seraient considérées comme plausibles dans un contexte d'enquête en personne ou par téléphone (West et Olson 2010). Les variations connues des composantes de la variance entre les groupes sont donc de 1 pour la variable normale, et de 0,1 pour la variable de Bernoulli.

Compte tenu de ces valeurs connues pour les composantes de l'écart intervieweurs dans la population hypothétique, nous avons appliqué chaque méthode décrite dans la section 2 [au moyen de valeurs uniformes diffuses et non informatives pour les composantes de la variance, selon les recommandations de Gelman (2006, section 7)] pour chaque échantillon hypothétique. Nous avons calculé les mesures empiriques suivantes à des fins de comparaison : 1) le biais empirique et relatif de l'estimateur; 2) l'EQM empirique de l'estimateur; 3) la couverture « fréquentiste » des intervalles de type Wald à 95 % (lorsque l'on utilise l'approche fréquentiste) et les ensembles crédibles à 95 % (lorsque l'on utilise l'approche bayésienne); et 4) les largeurs moyennes des intervalles de type Wald à 95 % et les ensembles crédibles. Le nombre d'intervalles de type Wald ne pouvant pas être calculés en raison des composantes estimées de la variance de 0 (sans erreur type connexe) a également été enregistré dans chaque cas. Toutes les simulations ont été effectuées au moyen de SAS, R et BUGS, et le code de simulation est disponible sur demande.

Le tableau 3.1 présente les résultats de l'étude de simulation. Les résultats suggèrent que pour les échantillons modérés à grands d'intervieweurs et de répondants, les deux approches donnent des estimateurs de la différence entre les composantes de la variance qui ont un biais relativement petit, comme prévu. On a découvert que l'approche fréquentiste donne des estimateurs ayant des valeurs empiriques plus petites de l'EQM. Ce n'est pas vraiment étonnant, étant donné la variabilité supplémentaire des estimations bayésiennes introduites par la prise en compte de l'incertitude des répartitions a priori des paramètres dans les distributions a priori non informatives. L'utilisation des distributions a priori informatives pourrait améliorer l'efficacité des estimations bayésiennes. Dans le

contexte d'un gros échantillon, les intervalles de confiance de 95 % et les ensembles crédibles calculés pour la différence entre les composantes de la variance semblent avoir des propriétés de couverture acceptables, l'approche bayésienne ayant un léger sous-dénombrement.

Tableau 3.1
Résultats de l'étude de simulation comparant les propriétés empiriques des approches fréquentistes et bayésiennes pour faire une inférence au sujet des différences entre les composantes de l'écart intervieweurs.

Tailles d'échantillon		Approche fréquentiste	Approche bayésienne
		Y Normale	
50 intervieweurs / groupe 50 cas / intervieweur (n = 2 500 / groupe)	Biais empirique	-0,0498	-0,0189
	Biais relatif	-4,98%	-1,89%
	EQM empirique	0,6546	0,8134
	IC/EC de 95%	0,960	0,920
	Largeur moyenne IC/EC 95%	3,1689	3,6283
	% des IC de Wald invalides	0,0%	--
			Y Bernoulli
20 intervieweurs / groupe 10 cas / intervieweur (n = 200 / groupe)	Biais empirique	-0,0020	-0,0046
	Biais relatif	-2,0%	-4,6%
	EQM empirique	0,0029	0,0033
	IC/EC de 95%	0,938	0,940
	Largeur moyenne IC/EC 95%	0,2142	0,2372
	% des IC de Wald invalides	11,5%	--
			Y Normale
20 intervieweurs / groupe 10 cas / intervieweur (n = 200 / groupe)	Biais empirique	-0,2341	-0,3508
	Biais relatif	-23,41%	-35,08%
	EQM empirique	6,9873	6,2869
	IC/EC de 95%	1,000	0,995
	Largeur moyenne IC/EC 95%	16,6313	18,3574
	% des IC de Wald invalides	54,0%	--
			Y Bernoulli
20 intervieweurs / groupe 10 cas / intervieweur (n = 200 / groupe)	Biais empirique	-0,0348	-0,0196
	Biais relatif	-34,8%	-19,6%
	EQM empirique	0,0345	0,0861
	IC/EC de 95%	1,000	0,980
	Largeur moyenne IC/EC 95%	1,2604	1,7970
	% des IC de Wald invalides	65,5%	--

Fait intéressant, 11,5 % des intervalles de confiance de 95 % de type Wald n'ont pas pu être calculés pendant l'analyse du résultat binaire des gros échantillons, parce qu'une des composantes de la variance estimée était égale à zéro (sans erreur type). Ce taux de « rejet à la vérification » pour les intervalles de Wald s'aggravait beaucoup pour les deux variables des petits échantillons, où les deux méthodes

produisaient également des estimations ayant un biais négatif. L'approche fréquentiste peut donc fournir une estimation de la différence et des intervalles de confiance connexes qui fonctionnent bien dans les gros échantillons avec des variables réparties normalement, mais dans les petits échantillons ou même les échantillons de taille modérée à grande ayant des variables non normales, les intervalles simples de type Wald pouvant être calculés au moyen des logiciels standard peuvent échouer une bonne partie du temps. Cet échec est attribuable au fait que la matrice de Hessian n'est pas inversible lorsqu'une composante de variance estimée est établie à zéro (c.-à-d. que la vraisemblance ne peut pas être estimée par un quadratique). Dans l'ensemble, ces résultats de simulation suggèrent donc que : 1) les deux approches auront un rendement semblable lorsqu'elles seront appliquées à des données d'enquête réelles avec des échantillons modérés à gros d'intervieweurs et de répondants; 2) l'approche bayésienne pourrait être la meilleure option si les intervalles (ou ensembles crédibles) pour la différence sont souhaités; et 3) il faut faire preuve de prudence lorsque l'on applique l'une ou l'autre des méthodes aux échantillons relativement petits d'intervieweurs et de répondants.

4 Application : La National Survey of Family Growth (NSFG) des États-Unis

Nous appliquons maintenant les approches fréquentistes et bayésiennes aux données d'enquête réelles recueillies au septième cycle de la NSFG (de juin 2006 à juin 2010). Le plan original de ce cycle de la NSFG (Groves et coll. 2009) a nécessité 16 trimestres de collecte des données d'un échantillon continu qui était représentatif à l'échelle nationale lorsqu'il a été terminé en juin 2010. Les données analysées dans ce document ont été recueillies à partir d'un échantillon national de 11 609 femmes de 15 à 44 ans, par 87 intervieweuses (aux tailles d'échantillons variables pour chaque intervieweuse). Pour obtenir plus de détails sur le plan et le fonctionnement du septième cycle de la NSFG, voir Lepkowski et coll. (2010) ou Groves et coll. (2009).

Pour chacune des 87 intervieweuses, l'information est disponible concernant l'âge (47,1 % ont 55 ans ou plus), les années d'expérience (43,7 % ont au moins cinq ans d'expérience), le nombre d'enfants (33,3 % ont deux enfants ou plus), l'état matrimonial (19,5 % n'ont jamais été mariées), les autres emplois (46,0 % avaient un autre emploi), les études postsecondaires (57,5 % ont obtenu un diplôme d'un programme collégial de quatre ans), l'expérience antérieure dans le cadre de la NSFG (82,8 % avaient travaillé aux cycles précédents) et l'appartenance ethnique (81,6 % sont blanches). Ces caractéristiques observables au niveau de l'intervieweuse serviront à diviser les intervieweuses en deux groupes (en l'absence d'une expérience randomisée idéale, comme décrit à la section 2).

Pour chacune des 11 609 répondantes, on mesure la parité (ou nombre de naissances vivantes) et un indicateur de l'activité sexuelle actuelle (indiquée par au moins un partenaire masculin actuel ou au moins un partenaire masculin au cours des 12 derniers mois) en vue d'une analyse. Bien que ces mesures semblent simples, les concepts mesurés peuvent être communiqués différemment par différentes intervieweuses (ce qui donne lieu à l'écart intervieweuses). La principale question analytique vise à

déterminer si ces différents groupes d'intervieweuses ont des composantes de la variance considérablement différentes pour ces variables d'enquête en particulier.

Nous examinons d'abord un MLGH pour la variable de la parité. Supposons que Y soit une variable aléatoire de Poisson avec un paramètre λ . Nous autorisons la surdispersion (ou dispersion extra-Poisson) dans Y , ce qui est très fréquent dans les variables de dénombrement (par exemple, la parité moyenne pour l'échantillon de 11 609 femmes est de 1,19, et la variance des valeurs mesurées de la parité est de 1,99). En nous appuyant sur Hilbe (2007) et Durham, Pardoe et Vega (2004), nous supposons que $\lambda = r\mu$, où r est une variable aléatoire $\text{GAMMA}(\alpha^{-1}, \alpha^{-1})$. On constate alors que Y a une répartition binomiale négative avec une moyenne μ et un paramètre d'échelle α :

$$\begin{aligned} E(Y) &= E(\lambda) = E(r\mu) = \mu E(r) = \mu \\ \text{var}(Y) &= E(\lambda) + \text{var}(\lambda) = E(r\mu) + \text{var}(r\mu) = \mu E(r) + \mu^2 \text{var}(r) = \mu(1 + \alpha\mu) \end{aligned}$$

Nous précisons un MLGH pour la valeur observée de la parité pour la répondante j interviewée par l'intervieweuse i , y_{ij} , comme suit :

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\lambda_i), \quad \lambda_i = r_i \mu_i \\ r_i &\sim \text{Gamma}(\alpha^{-1}, \alpha^{-1}) \\ \log(\mu_i) &= \beta_0 + \beta_1 I(\text{Groupe} = 1)_i + u_{i(1)} I(\text{Groupe} = 1)_i + u_{i(2)} I(\text{Groupe} = 2)_i \\ u_{i(1)} &\sim N(0, \tau_1^2), \quad u_{i(2)} \sim N(0, \tau_2^2). \end{aligned} \tag{4.1}$$

Dans ce modèle multiniveau de régression binomiale négative, $\exp(\beta_0)$ représente la parité prévue pour le groupe 2, $\exp(\beta_1)$ représente le changement multiplicatif prévu pour le groupe 1 par rapport au groupe 2, $u_{i(1)}$ est un effet aléatoire associé à l'intervieweuse i dans le groupe 1, et $u_{i(2)}$ est un effet aléatoire associé à l'intervieweuse i du groupe 2.

Ensuite, nous examinons un MLGH pour l'indicateur binaire de l'activité sexuelle actuelle. Supposons que $z_{ij} = 1$ si une répondante j indique l'activité sexuelle actuelle à l'intervieweuse i , et 0 autrement.

Nous précisons le modèle suivant pour cet indicateur binaire :

$$\begin{aligned} z_{ij} &\sim \text{Bernoulli}(p_i) \\ \ln\left[\frac{p_i}{1-p_i}\right] &= \beta_0 + \beta_1 I(\text{Groupe} = 1)_i + u_{i(1)} I(\text{Groupe} = 1)_i + u_{i(2)} I(\text{Groupe} = 2)_i \\ u_{i(1)} &\sim N(0, \tau_1^2), \quad u_{i(2)} \sim N(0, \tau_2^2). \end{aligned} \tag{4.2}$$

Dans ce modèle, $\exp(\beta_0)$ représente les probabilités prévues d'activité sexuelle actuelle pour le groupe 2, $\exp(\beta_1)$ représente la variation multiplicative prévue des probabilités d'activité sexuelle actuelle pour le groupe 1 par rapport au groupe 2, $u_{i(1)}$ est un effet aléatoire associé à l'intervieweuse i dans le groupe 1, et $u_{i(2)}$ est un effet aléatoire associé à l'intervieweuse i dans le groupe 2.

Nous intégrons les modèles (4.1) et (4.2) au moyen des deux approches décrites à la section 2. Pour l'approche fréquentiste, d'après les recommandations de la documentation mentionnée à la section 2, nous avons estimé les paramètres dans ces modèles au moyen de l'estimation de la pseudo-vraisemblance résiduelle (PVR), telle que mise en œuvre dans la procédure GLIMMIX du logiciel SAS/STAT. Toutes les analyses fréquentistes présentées dans cette section ont été répétées au moyen d'une quadrature adaptative pour estimer les fonctions de vraisemblance, et les principaux résultats n'ont pas changé; de plus, l'utilisation de la quadrature adaptative a donné des délais d'estimation plus longs.

Pour l'approche bayésienne, les répartitions a priori non informatives suivantes pour ces paramètres ont été utilisées. Ces répartitions a priori ont été sélectionnées en fonction d'une combinaison d'estimations à partir de l'intégration initiale du modèle naïf, ainsi que des recommandations de Gelman et Hill (2007) et de Gelman (2006, section 7) pour les répartitions a priori correctes mais non informatives pour les paramètres de la variance dans les modèles hiérarchiques ayant un nombre raisonnablement élevé de groupes (c.-à-d. plus de cinq) ou d'intervieweuses, dans le contexte actuel :

$$\begin{aligned}\beta_0 &\sim N(0,100) & \beta_1 &\sim N(0,100) \\ \tau_1^2 &\sim \text{Uniforme}(0,10) & \tau_2^2 &\sim \text{Uniforme}(0,10) \\ \ln(\alpha) &\sim N(0,100).\end{aligned}$$

Les valeurs a priori non informatives pour les effets fixes et le paramètre d'échelle (après transformation par le logarithme naturel) pour la variable du dénombrement binomial négatif (parité) indiquent qu'on s'attend à ce que ces paramètres soient quelque part dans la plage (-10, 10), tandis que les valeurs a priori non informatives pour les composantes de la variance sont des répartitions uniformes dans la plage (0, 10). Étant donné les estimations naïves initiales des effets fixes de -1 à 1 et les estimations initiales du paramètre d'échelle (non transformé) et des composantes de la variance de 0 à 5, ces valeurs a priori sont relativement diffuses, exprimant peu de connaissances antérieures au sujet de ces paramètres et laissant les données disponibles de la NSFG fournir la majeure partie de l'information. Les études antérieures comparant les composantes de l'écart intervieweurs pour des variables semblables sur le dénombrement pourraient également être utilisées dans les applications générales de cette technique afin de préciser des répartitions a priori plus informatives. Il est également important de souligner que le logiciel BUGS utilise des variances inversées pour la répartition normale, ce qui veut dire que 0,01 et les inverses des composantes de la variance seront indiqués dans les fonctions de la répartition normale (par exemple, le code WinBUGS utilisé pour les analyses est disponible à l'annexe).

Le tableau 4.1 présente des statistiques descriptives pour les intervieweuses dans chacun des groupes définis par les huit caractéristiques au niveau de l'intervieweuse. Ces statistiques descriptives comprennent le nombre d'intervieweuses dans chaque groupe (sur 87 au total), ainsi que la moyenne, l'écart-type (Éc.T.) et la plage pour le nombre de cas (tailles d'échantillons) affectés à chaque intervieweuse.

Tableau 4.1
Statistiques descriptives pour les intervieweuses de la NSFG dans chaque groupe défini par les huit caractéristiques au niveau de l'intervieweuse

	Nombre d'intervieweuses	Taille d'échantillon totale	Taille d'échantillon moyenne	Éc.T. des tailles d'échantillon	Étendue des tailles d'échantillon
Âge (Années)					
< 54	46	5 888	128,00	113,29	(18, 554)
55 et +	41	5 721	139,54	132,67	(12, 532)
Expérience					
< 5 ans	49	6 062	123,71	126,65	(12, 554)
5 ans et +	38	5 547	145,97	116,71	(18, 507)
Nbr. d'enfants					
< 2	58	7 756	133,72	113,28	(18, 532)
2 et +	29	3 853	132,86	140,53	(12, 554)
Déjà mariée					
Oui	70	9 923	141,76	129,00	(17, 554)
Non	17	1 686	99,18	83,49	(12, 377)
Autre emploi					
Non	47	5 406	115,02	95,49	(12, 532)
Oui	40	6 203	155,08	145,92	(17, 554)
Diplôme collégial					
Non	37	4 528	122,38	87,97	(18, 409)
Oui	50	7 081	141,62	142,71	(12, 554)
NSFG avant					
Non	15	1 155	77,00	39,17	(20, 166)
Oui	72	10 454	145,19	130,29	(12, 554)
Ethnicité					
Autre	16	1 781	111,31	75,53	(20, 297)
Blanche	71	9 828	138,42	130,35	(12, 554)

Les statistiques descriptives dans le tableau 4.1 indiquent une variance considérable des tailles des échantillons attribuées aux intervieweuses. Une approche de modélisation traitant les effets des intervieweurs comme des valeurs fixes ne conviendrait probablement pas à ces données, compte tenu des petites tailles pour certaines des intervieweuses (ce qui pourrait donner lieu à des estimations instables pour certaines intervieweuses). Au lieu de cela, une approche de modélisation empruntant de l'information à plusieurs intervieweuses (traitant les effets des intervieweuses comme des valeurs aléatoires) donnerait des estimations plus stables des moyennes pour chaque intervieweuse. Nous constatons également que pour trois des caractéristiques observables des intervieweuses (déjà mariée, expérience relative à la NSFG et ethnicité), un des deux groupes a moins de 20 intervieweuses, ce qui n'est pas idéal pour une estimation fiable des composantes de la variance (Hox 1998). Compte tenu des résultats de la simulation pour les

petits échantillons (section 3), nous tenons compte des effets des groupes sociodémographiques dans nos analyses.

De simples examens des répartitions des moyennes des mesures de la parité observée pour les intervieweuses dans chaque groupe sont présentés à la figure 4.1 ci-après, afin d'obtenir une première impression de l'ampleur de l'écart intervieweurs dans chaque groupe. La figure 4.1 présente des boîtes à pattes côte à côte des moyennes des intervieweuses pour la variable de la parité pour chaque groupe, la moyenne étant pondérée par l'attribution des tailles des échantillons, ainsi que la répartition globale des 11 609 mesures de la parité dans l'ensemble de données complet.

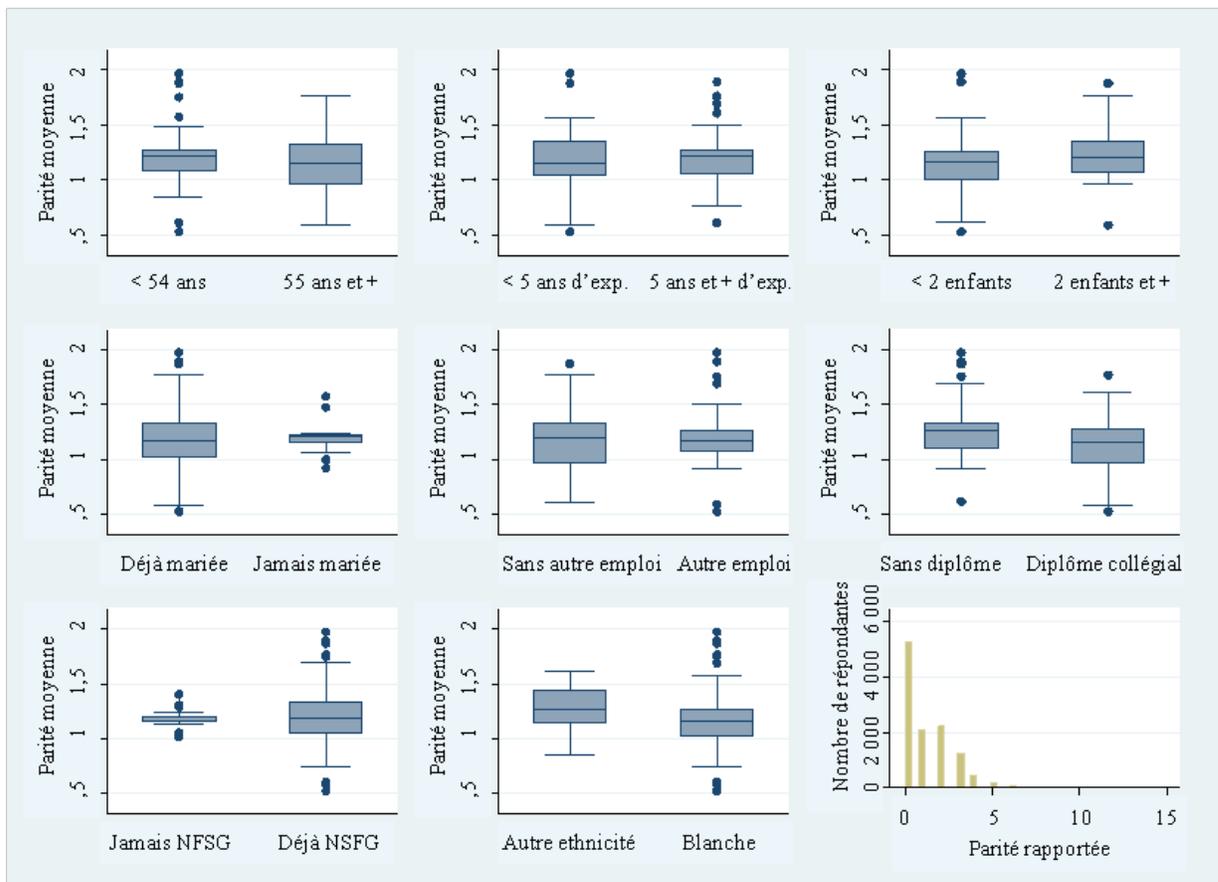


Figure 4.1 Répartitions des moyennes observées pour les intervieweuses de chaque groupe, les moyennes des intervieweuses étant pondérées par l'attribution de la taille de l'échantillon, ainsi que la répartition globale des mesures déclarées de la parité.

Les répartitions des moyennes des valeurs mesurées de la parité pour les intervieweuses dans la figure 4.1 donnent une première impression des groupes qui ont tendance à diverger pour ce qui est des composantes de l'écart intervieweurs. Le groupe d'intervieweuses qui n'ont jamais été mariées semble avoir une variance réduite, à l'instar du groupe d'intervieweuses sans expérience préalable relative à la NSFG. Les boîtes à pattes suggèrent également que les groupes ne varient pas beaucoup en ce qui concerne les moyennes de la parité, ce qui est rassurant (c.-à-d. que des groupes différents d'intervieweuses ne produisent pas de moyennes marginales différentes pour l'estimation d'intérêt). Enfin,

la répartition des valeurs observées de la parité pour la totalité des 11 609 répondantes a l'apparence prévue pour une variable mesurant certains événements relativement rares (naissances vivantes), avec une moyenne de 1,19 et une variance de 1,99.

Nous considérons ensuite les répartitions des proportions de femmes indiquant l'activité sexuelle actuelle parmi les intervieweuses dans chaque groupe (figure 4.2).

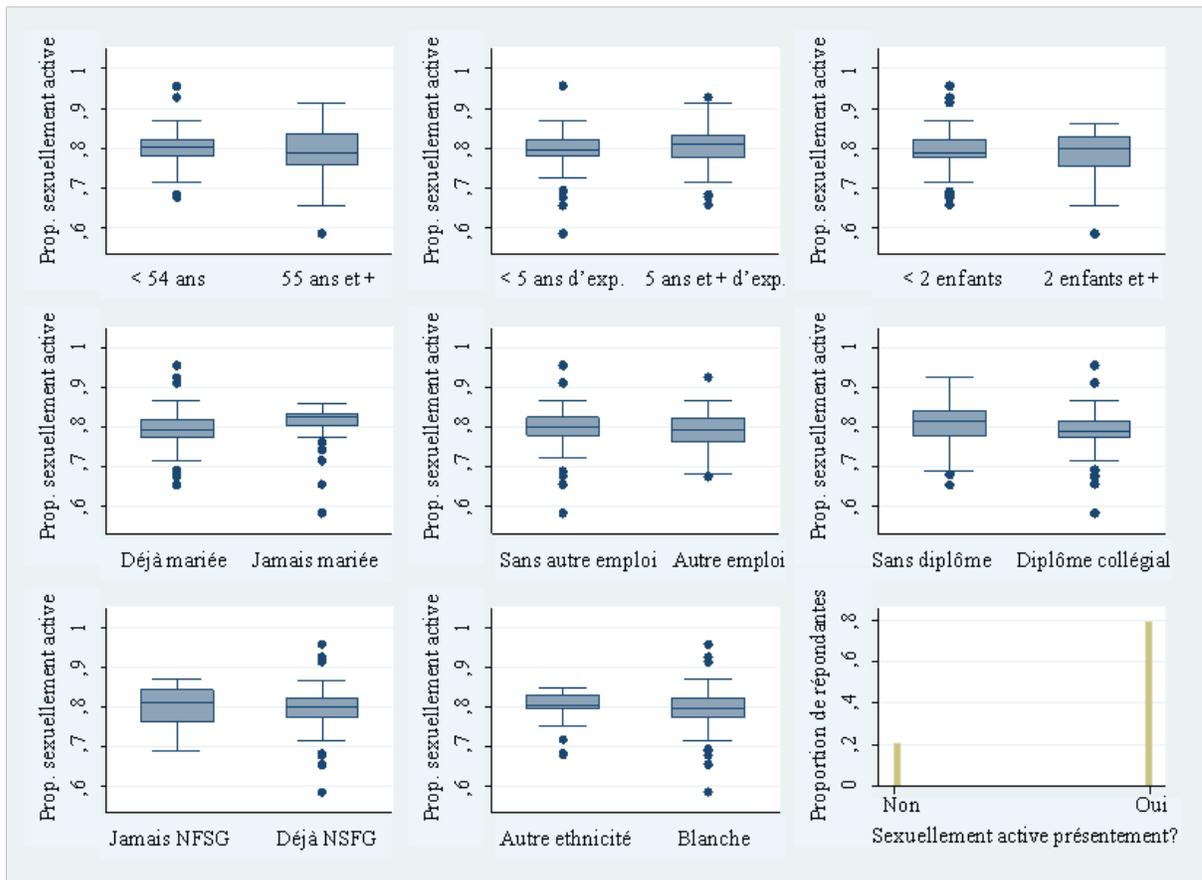


Figure 4.2 Répartitions des proportions observées de répondantes indiquant l'activité sexuelle actuelle pour les intervieweuses dans chaque groupe, les moyennes des intervieweuses étant pondérées par l'attribution de la taille de l'échantillon, ainsi que la répartition globale de l'indicateur d'activité sexuelle.

Nous voyons moins d'indices de variations de l'écart intervieweurs entre les groupes en général pour cette proportion, par rapport à la parité moyenne. Environ 80 % des répondantes ont indiqué qu'elles étaient actuellement actives sexuellement.

Le tableau 4.2 présente des estimations des paramètres dans chaque modèle binomial négatif pour la variable de la parité mesurée en fonction des deux approches analytiques. Ce tableau présente également les résultats des tests du rapport des vraisemblances comparant les deux composantes de l'écart intervieweurs (pour chaque paire de groupes) dans le contexte de l'approche fréquentiste, et ensembles crédibles de 95 % pour la différence entre les deux composantes de la variance dans le contexte de l'approche bayésienne.

Tableau 4.2
Estimations des paramètres des modèles de régression binomiale négative pour la parité et les comparaisons des composantes de l'écart intervieweurs dans le contexte des approches analytiques de rechange fréquentistes et bayésiennes.

Groupe de variables de l'intervieweuse	Approche fréquentiste (SAS PROC GLIMMIX)				Approche bayésienne (WinBUGS)			
	$\hat{\beta}_0$ (ET)/	$\hat{\alpha}$	$\hat{\tau}_1^2$ (ET)/	Test du rapport de vrais. : $\tau_1^2 = \tau_2^2$	$\hat{\beta}_0$ (Éc.T.)/	$\hat{\alpha}$	$\hat{\tau}_1^2$ (Éc.T.)/	95% EC : $\tau_1^2 - \tau_2^2$
	$\hat{\beta}_1$ (ET)	(ET)	$\hat{\tau}_2^2$ (ET)		$\hat{\beta}_1$ (Éc.T.)	(Éc.T.)	$\hat{\tau}_2^2$ (Éc.T.)	
Âge (1 = < 54 ans, 2 = 55 ans et +)	0,185(0,031)/ -0,007(0,043)	0,538 (0,018)	0,026(0,009)/ 0,024(0,008)	$\chi_1^2 = 0,03$, p = 0,873	0,183(0,033)/ -0,003(0,046)	0,685 (0,024)	0,025(0,010)/ 0,024(0,009)	(-0,026; 0,028)
Expérience (1 = < 5 ans, 2 = 5 ans et +)	0,201(0,033)/ -0,036(0,044)	0,537 (0,018)	0,024(0,008)/ 0,027(0,010)	$\chi_1^2 = 0,04$, p = 0,835	0,197(0,034)/ -0,031(0,045)	0,694 (0,027)	0,024(0,009)/ 0,027(0,011)	(-0,032; 0,024)
Nbr. d'enfants (1 = < 2, 2 = 2 et +)	0,254(0,036)/ -0,109(0,044)	0,537 (0,018)	0,023(0,007)/ 0,022(0,009)	$\chi_1^2 = 0,01$, p = 0,926	0,253(0,038)/ -0,109(0,045)	0,692 (0,025)	0,023(0,007)/ 0,023(0,012)	(-0,032; 0,024)
Déjà mariée (1 = Oui, 2 = Non)	0,184(0,029)/ -0,001(0,039)	0,537 (0,018)	0,030(0,008)/ 0,000(S.O.)*	$\chi_1^2 = 5,41$, p = 0,020	0,181(0,037)/ 0,004(0,045)	0,694 (0,025)	0,030(0,008)/ 0,003 (0,007)	(0,002; 0,048)
Autre emploi (1 = Oui, 2 = Non)	0,186(0,031)/ -0,009(0,043)	0,538 (0,018)	0,022(0,009)/ 0,027(0,008)	$\chi_1^2 = 0,15$, p = 0,699	0,188(0,032)/ -0,010(0,044)	0,688 (0,025)	0,020(0,010)/ 0,028(0,010)	(-0,036; 0,021)
Diplôme collégial (1 = Oui, 2 = Non)	0,242(0,031)/ -0,108(0,042)	0,538 (0,018)	0,023(0,008)/ 0,022(0,008)	$\chi_1^2 < 0,01$, p = 0,963	0,240(0,032)/ -0,106(0,044)	0,693 (0,024)	0,024(0,009)/ 0,021(0,010)	(-0,025; 0,030)
NSFG avant (1 = Oui, 2 = Non)	0,174(0,035)/ 0,010(0,043)	0,537 (0,018)	0,031(0,008)/ 0,000(S.O.)*	$\chi_1^2 = 8,26$, p = 0,004	0,169(0,036)/ 0,013(0,045)	0,692 (0,026)	0,030(0,008)/ 0,001(0,005)	(0,006; 0,050)
Ethnicité (1 = Blanche, 2 = Autre)	0,217(0,046)/ -0,044(0,052)	0,537 (0,018)	0,027(0,007)/ 0,018(0,011)	$\chi_1^2 = 0,38$, p = 0,536	0,220(0,051)/ -0,050(0,058)	0,690 (0,025)	0,026(0,008)/ 0,020(0,017)	(-0,045; 0,027)

* PROC GLIMMIX indiquait que la matrice estimée de variance-covariance des effets aléatoires n'était pas définie positive, et que l'estimation était établie à zéro à cause que l'estimation de la PVR de la composante de variance était négative. Le même résultat s'est produit lorsque l'on a utilisé la quadrature adaptative au lieu de la PVR.

Notes : Les estimations suivant l'approche bayésienne sont des médianes des tirages de répartitions a posteriori. ET = erreur-type asymptotique. Éc.T. = écart type des tirages de la répartition a posteriori. EC = ensemble crédible.

Conformément à notre étude de simulation à la section 3, les résultats dans le tableau 4.2 indiquent qu'il n'est pas rare que l'approche fréquentiste donne des estimations négatives des composantes de l'écart intervieweurs (qui amène SAS PROC GLIMMIX à établir les estimations égales à zéro, et à ne pas déclarer les erreurs types estimées pour les estimations), en particulier pour les groupes ayant de petits

échantillons d'intervieweurs. Dans deux cas, il s'ensuit une importante statistique sur le test du rapport des vraisemblances, qui pourrait suggérer que les deux composantes de la variance sont différentes. En revanche, l'approche bayésienne produit de très petites estimations des composantes de la variance, et un ensemble crédible de 95 % pour la différence entre les composantes de la variance. Par exemple, dans les cas de l'état matrimonial et de l'expérience préalable relative à la NSFG, nous voyons des estimations qui sont conformes à la figure 4.1, ce qui porte à croire qu'il y a une variance considérablement plus faible dans les mesures de la parité dans le groupe d'intervieweuses jamais mariées et le groupe d'intervieweuses sans expérience. Les ensembles crédibles pour les différences dans ces deux cas cadrent avec les tests fréquentistes, mais les limites inférieures de ces ensembles sont très proches de zéro, ce qui suggère que les différences, bien qu'importantes, ne sont peut-être pas très marquées. Nous considérons qu'il s'agit d'un avantage de l'approche bayésienne.

L'approche bayésienne donne des erreurs-types un peu plus importantes (ou des écarts-types a posteriori) pour les estimations des paramètres dans presque tous les cas, ce qui indique l'incertitude des estimations des paramètres qui est prise en compte par les répartitions a priori. L'utilisation des distributions a priori non informatives dans ce cas, qui pourrait entraîner une répartition a posteriori dominée par la fonction de vraisemblance, est la raison probable de la similarité dans ces mesures de l'incertitude, et des distributions a priori plus informatives pourraient augmenter l'efficacité des estimations bayésiennes. Les estimations des paramètres individuels et les inférences correspondantes connexes sont généralement très semblables lorsque l'on suit les deux approches, comme le suggère la documentation à la section 2, et les effets fixes estimés suggèrent que les groupes différents d'intervieweuses n'ont pas tendance à recueillir des mesures différentes pour la variable de la parité. Fait intéressant, les deux approches conviennent que les intervieweuses ayant moins d'enfants et/ou un diplôme collégial de quatre ans ont tendance à recueillir des mesures plus faibles pour la variable de la parité, mais ces différences pourraient bien être attribuables à d'autres covariables non prises en compte dans ces analyses. Enfin, nous voyons des estimations légèrement différentes du paramètre d'échelle binomiale négative selon les deux approches. Il fallait s'y attendre, puisque l'approche bayésienne utilise les médianes des répartitions a posteriori, tandis que l'approche fréquentiste utilise les modes des fonctions de vraisemblance. De plus, les répartitions a posteriori ne sont pas exactement égales aux fonctions de vraisemblance lorsque les distributions a priori appropriées sont utilisées. Les estimations fréquentistes du paramètre d'échelle étaient beaucoup plus proches des estimations bayésiennes lorsque l'on utilise la quadrature adaptative avec cinq points de quadrature pour estimer les vraisemblances binomiales négatives (résultats non montrés); les inférences fréquentistes pour les autres paramètres n'ont pas changé lorsqu'on utilisait cette méthode d'estimation de rechange.

Nous avons répété ces analyses pour l'indicateur d'activité sexuelle actuelle. Le tableau 4.3 présente les paramètres estimés des modèles de régression logistique multiniveaux suivant chacune des deux approches. Conformément à la figure 4.2, ces analyses ne révèlent aucun indice de différences entre les divers groupes d'intervieweuses pour les composantes de la variance ou les valeurs prévues de ce résultat. Les inférences étaient encore une fois très semblables lorsque l'on suivait les deux approches, et les variances des composantes de la variance estimée étaient encore une fois légèrement plus marquées selon l'approche bayésienne.

Tableau 4.3

Estimations des paramètres dans les modèles de régression logistique pour l'activité sexuelle actuelle et comparaisons des composantes de l'écart intervieweurs d'après les approches analytiques fréquentistes et bayésiennes de rechange.

Groupe de variables de l'intervieweuse	Approche fréquentiste (SAS PROC GLIMMIX)			Approche bayésienne (WinBUGS)		
	$\hat{\beta}_0$ (ET) / $\hat{\beta}_1$ (ET)	$\hat{\tau}_1^2$ (ET) / $\hat{\tau}_2^2$ (ET)	Test du rapport de vrais. : $\tau_1^2 = \tau_2^2$	$\hat{\beta}_0$ (Éc.T.) / $\hat{\beta}_1$ (Éc.T.)	$\hat{\tau}_1^2$ (ET) / $\hat{\tau}_2^2$ (ET)	95% EC: $\tau_1^2 - \tau_2^2$
Âge						
(1 = < 54 ans, 2 = 55 ans et +)	1,333 (0,066) / 0,032 (0,076)	0,008 (0,013) / 0,045 (0,024)	$\chi_1^2 = 2,05,$ $p = 0,153$	1,344 (0,055) / 0,024 (0,066)	0,009 (0,013) / 0,046 (0,028)	(-0,107; 0,016)
Expérience						
(1 = < 5 ans, 2 = 5 ans et +)	1,378 (0,064) / -0,050 (0,073)	0,004 (0,017) / 0,037 (0,020)	$\chi_1^2 = 1,52,$ $p = 0,217$	1,384 (0,051) / -0,061 (0,064)	0,005 (0,017) / 0,039 (0,023)	(-0,087; 0,024)
Nbr. d'enfants						
(1 = < 2, 2 = 2 et +)	1,362 (0,080) / -0,015 (0,088)	0,022 (0,015) / 0,033 (0,024)	$\chi_1^2 = 0,16,$ $p = 0,689$	1,363 (0,059) / -0,012 (0,070)	0,024 (0,016) / 0,037 (0,030)	(-0,094; 0,037)
Déjà mariée						
(1 = Oui, 2 = Non)	1,387 (0,130) / -0,045 (0,134)	0,020 (0,012) / 0,048 (0,041)	$\chi_1^2 = 0,58,$ $p = 0,447$	1,398 (0,090) / -0,053 (0,097)	0,021 (0,013) / 0,051 (0,055)	(-0,180; 0,035)
Autre emploi						
(1 = Oui, 2 = Non)	1,374 (0,043) / -0,046 (0,072)	0,026 (0,016) / 0,024 (0,020)	$\chi_1^2 = 0,01,$ $p = 0,927$	1,381 (0,045) / -0,051 (0,065)	0,029 (0,019) / 0,022 (0,022)	(-0,055; 0,063)
Diplôme collégial						
(1 = Oui, 2 = Non)	1,388 (0,051) / -0,063 (0,071)	0,016 (0,014) / 0,035 (0,022)	$\chi_1^2 = 0,60,$ $p = 0,439$	1,394 (0,052) / -0,072 (0,064)	0,014 (0,016) / 0,038 (0,024)	(-0,079; 0,033)
NSFG avant						
(1 = Oui, 2 = Non)	1,363 (0,103) / -0,012 (0,111)	0,020 (0,012) / 0,069 (0,055)	$\chi_1^2 = 1,20,$ $p = 0,273$	1,381 (0,113) / -0,024 (0,118)	0,021 (0,013) / 0,083 (0,084)	(-0,301; 0,019)
Ethnicité						
(1 = Blanche, 2 = Autre)	1,354 (0,077) / -0,004 (0,088)	0,024 (0,014) / 0,032 (0,031)	$\chi_1^2 = 0,05,$ $p = 0,816$	1,365 (0,080) / -0,013 (0,088)	0,025 (0,015) / 0,032 (0,044)	(-0,131; 0,044)

Notes : Les estimations suivant l'approche bayésienne sont des médianes de tirages de répartitions a posteriori. ET = erreur-type asymptotique. Éc.T. = écart-type des tirages de la répartition a posteriori. EC = ensemble crédible.

5 Mot de la fin

Le présent document a examiné les méthodes fréquentistes et bayésiennes pour comparer les composantes de l'écart intervieweurs pour les questions d'enquête non réparties normalement entre deux groupes indépendants d'intervieweurs d'enquête. Les méthodes sont basées sur une catégorie flexible de modèles linéaires généralisés hiérarchiques (les MLGH), qui permettent aux composantes de la variance pour deux groupes absolument exclusifs d'intervieweurs de varier, et des approches inférentielles de rechange basées sur ces modèles. Les résultats d'une étude de simulation suggèrent que les deux approches ont peu de biais empirique, des valeurs empiriques comparables de l'EQM et une bonne couverture pour des échantillons moyens ou gros d'intervieweurs et de répondants. Les analyses des données réelles de la National Survey of Family Growth (NSFG) des États-Unis suggèrent que les inférences basées sur les deux approches ont tendance à se ressembler. Nous constatons que le rendement similaire de ces deux approches est une bonne nouvelle pour les chercheurs d'enquête, en ce que les

fréquentistes et les bayésiens disposent des outils nécessaires pour analyser ce problème et tireront des conclusions similaires.

Quelques distinctions subtiles entre les deux approches sont ressorties des analyses, en particulier en ce qui concerne la taille des échantillons et les estimations des composantes de la variance qui sont extrêmement petites ou égales à zéro. Ces problèmes méritent une discussion approfondie, compte tenu de leurs conséquences pour la réalisation d'enquêtes. L'approche bayésienne illustrée ici est en mesure de tenir compte de l'incertitude de l'estimation des composantes de la variance pour former des ensembles crédibles et ne s'appuie pas sur la théorie asymptotique, mais nous avons constaté que les inférences au sujet des variations des composantes de la variance entre plusieurs sous-groupes différents d'intervieweurs de la NSFG (chacun d'une taille modérée) ne variaient pas par rapport à celles qui seraient faites au moyen des approches fréquentistes. Il faudrait approfondir l'analyse pour déterminer si nous obtiendrions ou non les mêmes résultats pour des groupes encore plus petits d'intervieweurs; l'étude de simulation présentée à la section 3 suggère qu'aucune des deux méthodes ne fonctionne bien dans un contexte où deux groupes de 20 intervieweurs recueillent des données auprès de 10 répondants chacun. Une application initiale de ces deux méthodes aux données à partir du premier trimestre de collecte des données pendant ce cycle de la NSFG (une vingtaine d'intervieweurs dans chacun des deux groupes interviewent une vingtaine de répondants chacun en moyenne) a donné des résultats semblables à ceux décrits ici pour les gros échantillons, certains indices portant à croire que l'approche bayésienne est plus conservatrice (West 2011).

En général, l'approche bayésienne offre une forme plus naturelle d'inférence pour ce problème, ce qui indique un éventail de valeurs pour la différence, où environ 95 % des différences se trouveront. Ces résultats pourraient intéresser certains consommateurs de produits d'une enquête donnée, au lieu de la simple valeur p pour un test du rapport des vraisemblances, qui ne donne pas aux utilisateurs une idée de l'ampleur des différences possibles. Dans le contexte fréquentiste, le test du rapport des vraisemblances pourrait être la seule méthode d'inférence disponible si l'estimation ponctuelle du maximum de la pseudo-vraisemblance pour une ou plusieurs des composantes de la variance est de zéro, sans erreur-type correspondante (empêchant le calcul d'intervalles de type Wald). Cette situation a été observée aussi bien dans les simulations que dans les analyses de la NSFG, en particulier pour les groupes ayant des petits échantillons d'intervieweurs; compte tenu de l'utilisation des tests du rapport des vraisemblances pour la théorie asymptotique, l'approche bayésienne pourrait être un meilleur choix pour les petits échantillons. Le rendement de l'approche bayésienne n'est toutefois pas idéal pour les très petits échantillons, comme le démontre l'étude de simulation à la section 3.

Nous avons observé deux importantes différences entre les sous-groupes d'intervieweurs pour les données de la NSFG, et dans chaque cas, le groupe ayant la plus petite variance avait une composante de variance estimée de zéro (sans erreur-type calculée) dans le contexte de l'approche fréquentiste. Les inférences obtenues en fonction de ces estimations (où les valeurs de vraisemblances ont été calculées au moyen des estimations de zéro pour les sous-groupes en question pendant les tests du rapport des vraisemblances) correspondaient à l'approche bayésienne. Nous rappelons aux lecteurs qui utilisent les méthodes fréquentistes que les petits échantillons d'intervieweurs ou les très petites quantités de variance parmi les intervieweurs pour des variables en particulier peuvent entraîner des estimations du maximum de vraisemblance négatives des composantes de la variance, ce qui peut être problématique pour l'interprétation de l'écart intervieweurs pour les groupes individuels. Certaines procédures logicielles

capables d'intégrer des modèles multiniveaux (p. ex. la procédure `gllamm` dans Stata, ou la fonction `lmer()` dans R) limitent les composantes de la variance aux valeurs supérieures à zéro pendant l'estimation pour empêcher ce problème, ce qui peut prolonger les délais d'estimation. D'autres procédures logicielles (comme `GLIMMIX` dans SAS) fixeront tout simplement ces estimations négatives à zéro, et échoueront à calculer une erreur-type. Bien que ces composantes de la variance ne puissent pas en principe être égales à zéro, nous suggérons d'interpréter ces résultats pour prouver qu'il y a une variance négligeable entre les intervieweurs d'un groupe donné. Bates (2009) s'oppose à l'utilisation d'erreurs-types pour faire des inférences au sujet des composantes de la variance dans le contexte fréquentiste, en particulier lorsque les composantes de la variance sont proches de zéro, au lieu de suggérer que la fonction de la déviance profilée devrait être utilisée pour visualiser la précision des estimations. Cette approche et l'approche de Wald pour calculer les intervalles de confiance seront toujours limitées aux petits échantillons.

Nous ne voyons pas de problème empirique à utiliser ces estimations zéro pour effectuer les tests du rapport des vraisemblances démontrés ici pour comparer des groupes d'intervieweurs, étant donné que les tirages bayésiens des composantes de la variance dans ces groupes seraient également très petits. Cependant, dans le cas de la variance de l'estimation de l'écart intervieweurs pour les groupes individuels, un examen de la sensibilité des inférences bayésiennes aux choix de différentes répartitions a priori pour les composantes de la variance devrait être effectué lorsque des composantes de la variance proches de zéro sont prévues, ou que le nombre d'intervieweurs est relativement faible (Browne et Draper 2006; Lambert et coll. 2005). De plus, si les chercheurs d'enquête souhaitent *prédire* les effets aléatoires des intervieweurs dans le cas où les composantes de l'écart intervieweurs devraient être proches de zéro, les méthodes fréquentistes et bayésiennes donnent toutes deux des résultats très médiocres, et la prédiction n'est pas recommandée dans ce cas (Singh et coll. 1998, p. 390). Voir Savalei et Kolenikov (2008) pour obtenir plus d'information sur la question de la variance zéro.

Cette étude n'était certainement pas sans limites. Nous reconnaissons que la conception de la NSFG, où les intervieweurs sont habituellement affectés à une seule zone d'échantillonnage principale, ne permettait pas une attribution imbriquée de cas échantillonnés aux intervieweurs. Par conséquent, il est difficile de démêler les effets des intervieweurs et les effets des zones d'échantillonnage principales. Les méthodes décrites dans le présent document peuvent facilement intégrer des covariables supplémentaires au niveau de l'intervieweur ou de la région, afin d'expliquer la variance parmi les intervieweurs ou les régions en raison des covariables observables. Il faudrait continuer la recherche en général pour déterminer comment estimer l'écart intervieweurs en présence d'un plan d'échantillonnage strictement non imbriqué, et nous n'avons pas abordé cette question ouverte dans le présent document. Comme mentionné à la section 1, des plans d'échantillonnage ont été utilisés dans les études récentes pour démêler les effets des intervieweurs des effets des régions. Les études ultérieures devraient étudier la capacité des deux approches examinées dans le présent document de détecter les différences entre les composantes de l'écart intervieweurs lorsque l'on utilise des modèles multiniveaux recoupsés qui comprennent aussi les effets des régions dans un plan d'échantillonnage imbriqué.

De même, nous n'avons pas tenu compte des caractéristiques d'échantillonnage complexe de la NSFG (c.-à-d. pondération ou échantillonnage en grappes stratifié) dans les analyses. La théorie à la base de l'estimation des paramètres dans les modèles multiniveaux en présence de poids d'enquête nécessite des poids pour les répondants et les grappes aux niveaux supérieurs, en l'occurrence les intervieweurs (Rabe-Hesketh et Skrondal 2006; Pfefferman, Skinner, Holmes, Goldstein et Rasbash 1998). Les analyses

présentées ici présumant effectivement que nous avons un échantillon d'intervieweurs pour une population plus grande qui a été sélectionnée avec une probabilité égale, et que tous les répondants pour chaque intervieweur avaient un poids égal. Les méthodes décrites par Gabler et Lahiri (2009) pourraient être utiles pour pallier cette faiblesse, et les analystes pourraient également inclure les effets fixes des poids d'enquête ou des codes de stratification dans les modèles proposés ici. Nous réservons ces approfondissements à des recherches futures.

Enfin, ce document n'a pas tenu compte d'un autre aspect riche de l'approche bayésienne, en ce que les tirages a posteriori des 87 effets aléatoires des intervieweurs dans les modèles ont également été générés par l'algorithme d'échantillonnage de Gibbs BUGS. Ces tirages permettraient aux gestionnaires d'enquête de faire des inférences au sujet des effets qu'ont certains intervieweurs sur certaines mesures d'enquête. Une mise à jour constante et régulière de ces répartitions a posteriori au fil du déroulement du cycle de collecte des données permettrait aux gestionnaires d'enquête d'intervenir lorsque les répartitions a posteriori pour des intervieweurs en particulier suggèrent que ces intervieweurs ont des effets non nuls sur les mesures d'enquête.

Remerciements

Les auteurs sont reconnaissants du soutien d'un contrat avec le National Center for Health Statistics, qui a permis la réalisation du septième cycle de la National Survey of Family Growth (contrat 200-2000-07001).

Annexe

A.1 Exemple de code

Voici un exemple de code pour intégrer les types de modèles abordés dans le document au moyen de SAS PROC GLIMMIX. Dans ce code, PARITY et SEXMAIN sont les variables de dénombrement et binaire respectivement, mesurées pour les répondants de la NSFG, FINAL_INT_ID est un code d'utilisateur final de l'intervieweur, et INT_NVMARRIED est une variable qui indique qu'un intervieweur n'a jamais été marié. L'option ASYCOV indique les estimations asymptotiques des variances et des covariances des composantes de la variance estimée.

```
/* état matrimonial */  
  
proc glimmix data = bayes.final_analysis asycov;  
  class final_int_id int_nvmarried;  
  model parity = int_nvmarried / dist = negbin link = log solution cl;  
  random int / subject = final_int_id group = int_nvmarried;  
  covtest homogeneity / cl (type = plr);  
  nloptions tech=nrridg;  
run;
```

```

proc glimmix data = bayes.final_analysis asycov;
  class final_int_id int_nvmarried;
  model sexmain (event = "1") = int_nvmarried / dist = binary link = logit
solution cl;
  random int / subject = final_int_id group = int_nvmarried;
  covtest homogeneity / cl (type = plr);
  nloptions tech=nrridg;
run;

```

Nous fournissons également un exemple de code WinBUGS pour intégrer les modèles au moyen des approches bayésiennes dont il est question ci-après. Nous appelons le code WinBUGS du logiciel R. Dans ce code, LOWAGE.G est un indicateur au niveau de l'intervieweur (avec 87 valeurs) d'appartenance à la tranche d'âge plus jeune des intervieweurs, et HIGHAGE.G est un indicateur d'appartenance au groupe plus âgé. Le code complet, y compris le code créant les variables utilisées ci-après, peut être obtenu des auteurs sur demande.

```

# Charger les progiciels nécessaires pour utiliser BUGS à partir de R

library(arm)
library(R2WinBUGS)

##### Analyses de parité

# Fichiers BUGS pour les groupes d'âge et la parité (age_nb.bug)

model {
  for (i in 1:n){
    parity[i] ~ dpois(lambda[i])
    lambda[i] <- rho[i]*mu[i]
    log(mu[i]) <- b0[intid[i]]
    rho[i]~dgamma(alpha,alpha)
  }

  for (j in 1:J){
    b0[j] ~ dnorm(b0.hat[j], tau.b0[highage.g[j]+1])
    b0.hat[j] <- beta0 + beta1*lowage.g[j]
  }

  beta0 ~ dnorm(0,0.01)
  beta1 ~ dnorm(0,0.01)
  alpha <- exp(logalpha)
  logalpha ~ dnorm(0,0.01)

  for (k in 1:2){
    tau.b0[k] <- pow(sigma.b0[k], -2)
    sigma.b0[k] ~ dunif(0,10)
  }
}

# Simulations pour le modèle parité/groupe d'âge dans BUGS

n <- length(parity)
J <- 87
age.data <- list("n", "J", "parity", "intid", "highage.g", "lowage.g")

```

```

age.inits <- fonction(){
  list (b0=rnorm(J), beta0=rnorm(1), beta1=rnorm(1), sigma.b0=runif(2),
  logalpha=rnorm(1))}
age.parameters <- c("b0", "beta0", "beta1", "sigma.b0", "alpha")
age.l <- bugs(age.data, age.inits, age.parameters, "age_nb.bug", n.chains = 3,
n.iter=5000, debug=TRUE,
bugs.directory="C:/Users/bwest/Desktop/winbugs14/WinBUGS14")

attach.bugs(age.l)

# Pour les tableaux de resultats et d'inférences

resultsmat <- cbind(numeric(6),numeric(6),numeric(6),numeric(6))

resultsmat[1,1] <- quantile(beta0,0.5)
resultsmat[1,2] <- sd(beta0)
resultsmat[1,3] <- quantile(beta0,0.025)
resultsmat[1,4] <- quantile(beta0,0.975)

resultsmat[2,1] <- quantile(beta1,0.5)
resultsmat[2,2] <- sd(beta1)
resultsmat[2,3] <- quantile(beta1,0.025)
resultsmat[2,4] <- quantile(beta1,0.975)

resultsmat[3,1] <- quantile(sigma.b0[,1]^2,0.5)
resultsmat[3,2] <- sd(sigma.b0[,1]^2)
resultsmat[3,3] <- quantile(sigma.b0[,1]^2,0.025)
resultsmat[3,4] <- quantile(sigma.b0[,1]^2,0.975)

resultsmat[4,1] <- quantile(sigma.b0[,2]^2,0.5)
resultsmat[4,2] <- sd(sigma.b0[,2]^2)
resultsmat[4,3] <- quantile(sigma.b0[,2]^2,0.025)
resultsmat[4,4] <- quantile(sigma.b0[,2]^2,0.975)

resultsmat[5,1] <- quantile(1/alpha,0.5)
resultsmat[5,2] <- sd(1/alpha)
resultsmat[5,3] <- quantile(1/alpha,0.025)
resultsmat[5,4] <- quantile(1/alpha,0.975)

vardiff <- sigma.b0[,1]^2 - sigma.b0[,2]^2
resultsmat[6,1] <- quantile(vardiff,0.5)
resultsmat[6,2] <- sd(vardiff)
resultsmat[6,3] <- quantile(vardiff,0.025)
resultsmat[6,4] <- quantile(vardiff,0.975)

resultsmat

##### Analyses de l'activité sexuelle actuelle

# Fichiers BUGS pour les groupes d'âge et l'activité sexuelle (age_bin.bug)

model {
  for (i in 1:n){
    sexmain[i] ~ dbern(p[i])
    logit(p[i]) <- b0[intid[i]]
  }
}

```

```

for (j in 1:J){
  b0[j] ~ dnorm(b0.hat[j], tau.b0[highage.g[j]+1])
  b0.hat[j] <- beta0 + beta1*lowage.g[j]
}
beta0 ~ dnorm(0,0.01)
beta1 ~ dnorm(0,0.01)

for (k in 1:2){
  tau.b0[k] <- pow(sigma.b0[k], -2)
  sigma.b0[k] ~ dunif(0,10)
}
}

# Simulations pour le modèle parité/groupe d'âge dans BUGS

n <- length(sexmain)
J <- 87
age.data <- list("n", "J", "sexmain", "intid", "highage.g", "lowage.g")
age.inits <- fonction(){
  list (b0=rnorm(J), beta0=rnorm(1), beta1=rnorm(1), sigma.b0=runif(2))}
age.parameters <- c("b0", "beta0", "beta1", "sigma.b0")
age.l <- bugs(age.data, age.inits, age.parameters, "age_bin.bug", n.chains =
3, n.iter=5000, debug=TRUE,
bugs.directory="C:/Users/bwest/Desktop/winbugs14/WinBUGS14")

attach.bugs(age.l)

```

Bibliographie

- Bates, D. (2009). Assessing the precision of estimates of variance components. *Presentation to the Max Planck Institute for Ornithology*, Seewiesen, 21 juillet 2009. La présentation peut être téléchargée à partir de <http://lme4.r-forge.r-project.org/slides/2009-07-21-Seewiesen/4PrecisionD.pdf>.
- Biemer, P.P. et Trewin, D. (1997). A review of measurement error effects on the analysis of survey data. Chapitre 27 de *Survey Measurement and Process Quality*, Editeurs Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz et Trewin. Wiley-Interscience, 603-632.
- Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.
- Browne, W.J. et Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3), 473-514.
- BUGS, <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.html>.
- Carlin, B.P. et Louis, T.A. (2009). *Bayesian Methods for Data Analysis*. Chapman and Hall / CRC Press.
- Chaloner, K. (1987). A Bayesian approach to the estimation of variance components for the unbalanced one-way random model. *Technometrics*, 29(3), 323-337.

- Collins, M. et Butcher, B. (1982). Interviewer and clustering effects in an attitude survey. *Journal of the Market Research Society*, 25, 39-58.
- Durham, C.A., Pardoe, I. et Vega, E. (2004). A methodology for evaluating how product characteristics impact choice in retail settings with many zero observations: An application to restaurant wine purchase. *Journal of Agricultural and Resource Economics*, 29(1), 112-131.
- Durrant, G.B., Groves, R.M., Staetsky, L. et Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, 74, 1-36.
- Faraway, J.J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall / CRC Press: Boca Raton, FL.
- Farrell, P.J. (2000). Bayesian inference for small area proportions. *Sankhya: The Indian Journal of Statistics, Series B (1960-2002)*, 62(3), 402-416.
- Fowler, F.J. et Mangione, T.W. (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: Sage.
- Gabler, S. et Lahiri, P. (2009). De la définition et de l'interprétation de la variabilité d'intervieweur pour un plan d'échantillonnage complexe. *Techniques d'enquête*, 35(1), 91-106.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515-533.
- Gelman, A., Carlin, J.B., Stern, H.S. et Rubin, D.B. (2004). *Bayesian Data Analysis*. Chapman and Hall / CRC Press.
- Gelman, A. et Hill, J. (2007). *Data Analysis using Regression and Multilevel / Hierarchical Models*. Cambridge University Press.
- Gelman, A. et Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.
- Gilks, W.R. et Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.
- Goldstein, H. (1995). *Multilevel Statistical Models, Second Edition*. Kendall's Library of Statistics 3, Edward Arnold.
- Groves, R.M. (2004). *Survey Errors and Survey Costs (2nd Edition)*. Dans le chapitre 8 : The Interviewer as a Source of Survey Measurement Error. Wiley-Interscience.
- Groves, R.M., Mosher, W.D., Lepkowski, J.M. et Kirgis, N.G. (2009). Planning and development of the continuous National Survey of Family Growth. National Center for Health Care Statistics. *Vital Health Statistics*, 1(48).
- Hilbe, J.M. (2007). *Negative Binomial Regression*. Cambridge University Press.
- Hox, J. (1998). *Multilevel Modeling: When and Why*. Dans I. Balderjahn, R. Mathar et M. Schader (Eds.). Classification, data analysis, and data highways. New York: Springer-Verlag, 147-154.

- Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, 57, 92-115.
- Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R. et Jones, D.R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, 24(15), 2401-2428.
- Lepkowski, J.M., Mosher, W.D., Davis, K.E., Groves, R.M. et Van Hoewyk, J. (2010). The 2006-2010 National Survey of Family Growth: sample design and analysis of a continuous survey. National Center for Health Statistics, *Vital and Health Statistics*, 2(150), juin 2010.
- Lynn, P., Kaminska, O. et Goldstein, H. (2011). Panel attrition: how important is it to keep the same interviewer? *ISER Working Paper Series*, Article 2011-02.
- Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, 109, 325-378.
- Mangione, T.W., Fowler, F.J. et Louis, T.A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics*, 8(3), 293-307.
- Molenberghs, G. et Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag, Berlin.
- O'Muircheartaigh, C. et Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society, Series A*, 161(1), 63-77.
- O'Muircheartaigh, C. et Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society, Series A*, 162(3), 437-446.
- Pfefferman, D., Skinner, C.J., Holmes, D.J., Goldstein, H. et Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B*, 60(1), 23-40.
- Pinheiro, J.C. et Chao, E.C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15, 58-81.
- Rabe-Hesketh, S. et Skrondal, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, Series A*, 169, 805-827.
- Raudenbush, S.W. et Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Newbury Park, CA.
- Rodriguez, G. et Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: a case-study. *Journal of the Royal Statistical Society, Series A*, 164(2), 339-355.
- SAS Institute, Inc. (2010). Documentation en ligne pour la procédure GLIMMIX.
- Savalei, V. et Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, 13(2), 150-170.

- Schaeffer, N.C., Dykema, J. et Maynard, D.W. (2010). *Handbook of Survey Research, Second Edition*. Dans *Interviewers and Interviewing*. J.D. Wright et P.V. Marsden (eds). Bingley, U.K.: Emerald Group Publishing Limited.
- Schnell, R. et Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics*, 21(3), 389-410.
- Schober, M. et Conrad, F. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 61, 576-602.
- Singh, A.C., Stukel, D.M. et Pfeiffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *Journal of the Royal Statistical Society, Series B*, 60(2), 377-396.
- Ugarte, M.D., Goicoa, T. et Militino, A.F. (2009). Empirical bayes and fully bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics and Data Analysis*, 53, 2938-2949.
- Van Tassell, C.P. et Van Vleck, L.D. (1996). Multiple-trait Gibbs sampler for animal models: Flexible programs for Bayesian and likelihood-based (co)variance component inference. *Journal of Animal Science*, 74, 2586-2597.
- Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, 26, 37-52.
- West, B.T. (2011). Bayesian analysis of between-group differences in variance components in hierarchical generalized linear models. Dans *JSM Proceedings, Survey Research Methods Section*. Alexandria, VA: *American Statistical Association*, 1828-1842.
- West, B.T. et Galecki, A.T. (2011). An overview of current software procedures for fitting linear mixed models. *The American Statistician*, 65(4), 274-282.
- West, B.T., Kreuter, F. et Jaenichen, U. (2013). "Interviewer" effects in face-to-face surveys: A function of sampling, measurement error, or nonresponse? *Journal of Official Statistics*, 29(2), 277-297.
- West, B.T. et Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly*, 74(5), 1004-1026.
- Zhang, D. et Lin, X. (2010). Variance component testing in generalized linear mixed models for longitudinal / clustered data and other related topics. *Random Effect and Latent Variable Model Selection. Springer Lecture Notes in Statistics*, Volume 192.