Techniques d'enquête, décembre 2002
Vol. 28, N° 2, pp. 171-184
Statistique Canada, N° 12-001-XPB au catalogue

Évaluation du biais lié à diverses stratégies de prise de contact dans les enquêtes téléphoniques sur l'emploi du temps

Jay Stewart [1]

Résumé

Dans le cas de la plupart des enquêtes téléphoniques sur l'emploi du temps, on appelle les répondants potentiels une journée donnée pour leur demander de déclarer leurs activités de la journée précédente. Comme la plupart ne sont pas disponibles la journée du premier appel, cette méthode d'enquête fait que la probabilité d'interviewer la personne au sujet d'une journée de référence particulière risque d'être corrélée aux activités qui ont eu lieu durant cette journée de référence. De surcroît, le biais de non-contact a plus d'importance dans le cas des enquêtes sur l'emploi du temps que dans les autres, parce que les réponses par procuration ne peuvent être acceptées. Par conséquent, il est essentiel, pour ces enquêtes, d'adopter une stratégie prévoyant des tentatives subséquentes de prise de contact avec les répondants. Une stratégie de prise de contact spécifie le calendrier des prises de contacts et la période de travail sur le terrain. Les auteurs de publications antérieures ont défini deux calendriers pour procéder à ces tentatives subséquentes : un calendrier basé sur les journées convenables pour le répondant et un calendrier basé sur des journées désignées. La plupart de ces auteurs recommandent d'adopter le calendrier à journées désignées, mais offrent peu de données pour étayer ce choix. Dans le présent article, nous utilisons des simulations informatiques pour examiner le biais associé au calendrier à journées convenables et à trois versions du calendrier à journées désignées. Les résultats indiquent qu'il est préférable d'utiliser le calendrier à journées désignées et valident les recommandations faites dans les publications antérieures. Le calendrier à journées convenables introduit un biais systématique, en ce sens que le temps consacré à des activités en-dehors du foyer ont tendance à être surestimées. Mais avant tout, les estimations fondées sur le calendrier à journées convenables sont sensibles à la variance de la probabilité d'un contact. Par contre, un calendrier à journées désignées avec possibilité de remise de l'interview ne crée qu'un biais très faible et produit des résultats robustes pour un large éventail d'hypothèses quant au profil des activités durant les diverses journées de la semaine.

Mots clés : Enquêtes téléphoniques sur l'emploi du temps; stratégies de prise de contact; biais; simulations informatiques.

1. Introduction

Les enquêtes téléphoniques sur l'emploi du temps posent un problème unique de collecte des données, parce qu'on appelle les répondants une journée particulière pour leur demander de déclarer leurs activités de la journée précédente. La difficulté tient au fait qu'on n'arrive pas à joindre la plupart des répondants – environ 75 % (Kalton 1985) – lors du premier appel, ce qui oblige à faire d'autres tentatives de prise de contact. Dans le cas de la plupart des enquêtes, le moment où ces tentatives supplémentaires sont faites n'importe pas, car on demande aux répondants de faire une déclaration concernant une période de référence fixe. En outre, dans la plupart des enquêtes, le problème de remémoration n'est pas trop important même si on reprend contact avec les répondants plusieurs jours après le premier appel. Par contre, dans le cas des enquêtes sur l'emploi du temps, la capacité qu'ont les répondants de se souvenir de leurs activités durant une journée particulière diminue considérablement après un jour ou deux, si bien qu'il faut attribuer une nouvelle journée de référence au répondant si on ne réussit pas à le joindre au moment du premier appel. Comme nous le verrons plus loin, dans de telles conditions, il se peut que la probabilité d'interviewer le répondant au sujet d'une journée de référence particulière soit reliée aux activités qui ont eu lieu durant cette journée de référence. Par conséquent, il est essentiel, pour ce genre d'enquête, d'élaborer une stratégie de prise subséquente de contact avec le répondant qui n'introduit pas de biais.

Stratégies de prise de contact

Une stratégie de prise de contact comprend un calendrier de prise de contact et une période de travail sur le terrain. Le calendrier de prise de contact spécifie quelles journées de la semaine les tentatives de contact seront faites et la période de travail sur le terrain précise le nombre maximum de semaines durant lesquelles des tentatives seront faites.

Les calendriers de prise de contact rentrent dans deux grandes catégories : les calendriers à journées désignées et les calendriers à journées convenables. Pour l'une et l'autre de ces catégories, on affecte aléatoirement chaque répondant à une journée initiale d'appel. S'il y a effectivement prise de contact lors de ce premier appel, l'intervieweur essaye de recueillir l'information au sujet de la journée de référence, qui est la journée précédant celle de l'appel. Par contre, à l'étape des tentatives subséquentes de prise de contact, les deux stratégies diffèrent.

Dans le cas d'un calendrier à journées désignées, deux démarches peuvent être adoptées pour faire les tentatives subséquentes de prise de contact. L'intervieweur pourrait

1. Jay Stewart, Office of Employment Research and Program Development, Bureau of Labor Statistics, 2 Massachusetts Avenue, NE Room 4945.

appeler le répondant à une date ultérieure et lui demander de déclarer ses activités lors de la journée originale de référence. Cette démarche permet de garder la journée originale de référence, mais allonge la période de remémoration. Harvey (1993) recommande de ne pas permettre que la période de remémoration soit supérieure à deux jours. La deuxième démarche consiste à remettre l'interview et à attribuer au répondant une nouvelle journée de référence. Kalton (1985) recommande de repousser l'interview d'une semaine exactement, de sorte que la nouvelle journée de référence corresponde au même jour de la semaine que la journée de référence originale.

Ces démarches ne sont pas mutuellement exclusives. Par exemple, le calendrier à journées désignées de Statistique Canada permet aux intervieweurs d'appeler les répondants jusqu'à deux jours après la journée de référence (Statistique Canada 1999) et de repousser l'interview d'une semaine si le répondant ne peut être rejoint après la deuxième journée d'essai. L'interview ne peut pas être remise de plus de trois semaines (Statistique Canada). Par exemple, si la journée de référence initiale est le lundi 1er, on appelle le répondant le mardi 2 et, au besoin, le mercredi 3. Si aucune interview n'a lieu l'une de ces journées, on appelle le répondant le mardi 9 et, au besoin, le mercredi 10, et on lui demande de déclarer ses activités du lundi 8. Ce processus se poursuit jusqu'à ce que le répondant soit interviewé, qu'il refuse de participer ou que quatre semaines se soient écoulées.

Le calendrier à journées convenables pour le répondant ne permet pas de garder la journée de référence désignée. Si aucun contact n'est pris, l'intervieweur appelle la journée suivante et chaque journée subséquente jusqu'à ce qu'il rejoigne le répondant. Lorsqu'il y a eu contact, l'intervieweur essaye de procéder à l'interview ou, si le répondant refuse de participer à l'interview à ce moment-là, fixe une autre journée convenable pour le répondant. La journée de référence est toujours la journée qui a précédé l'interview. Soulignons que puisque les répondants ne planifient vraisemblablement pas d'interview lors de journées occupées, leur permettre de choisir la journée de l'interview revient pratiquement au même que laisser à l'intervieweur la possibilité de proposer des journées consécutives (ou d'appeler pendant plusieurs journées consécutives) jusqu'à ce que le répondant potentiel accepte de répondre. Donc, on pourrait considérer le calendrier à journées convenables comme étant fonctionnellement identique à un calendrier avec tentatives de prise de contact chaque jour.

Une variante du calendrier à journées convenables décrit plus haut a été utilisée lors de l'Environmental Protection Agency (EPA) Time Diary Study de 1992–1994, réalisée par l'Université du Maryland (voir Triplett 1995). Plutôt que d'affecter les répondants à une journée d'appel initiale, on les a réparti en un échantillon des jours de semaine et un échantillon des fins de semaine. Par exemple, les répondants faisant partie de l'échantillon des fins de semaine pouvaient être appelés le dimanche (pour faire une déclaration concernant le samedi) ou le lundi (pour faire une déclaration concernant le dimanche). Les intervieweurs avaient pour instruction de faire au moins 20 tentatives d'appel avant de finaliser le cas comme étant non accompli.

Les auteurs de la plupart des articles de méthodologie appuient l'utilisation du calendrier à journées désignées (Kinsley et O'Donnell 1983; Kalton 1985; Lyberg 1989; Harvey 1993 et Harvey 1999). Par exemple, Lyberg (1989) soutient que le calendrier à journées convenables peut introduire un biais, car « le répondant peut choisir une journée où il n'est pas occupé, une journée où il ne s'adonne pas à un comportement socialement inacceptable, une journée qu'il considère comme représentative, *etc.* ». Kinsley et O'Donnell (1983) soutiennent, quant à eux, que l'horaire à journées convenables pourrait produire une exagération du nombre d'événements qui surviennent en-dehors de la maison, car le répondant est plus susceptible d'être interviewé durant une journée qui suit immédiatement une journée où il était hors de la maison.

Deux de ces études font une comparaison directe des calendriers à journées désignées et à journées convenables (Kinsley et O'Donnell 1983; Lyberg 1989). Dans Kinsley et O'Donnell (1983), le plan d'expérience consiste à diviser l'échantillon en deux groupes. Ces auteurs ont observé que les deux catégories de calendriers produisaient des taux de réponse similaires et que la composition démographique était la même pour les deux échantillons. Ils ont également observé que le temps estimatif passé en-dehors de la maison était nettement plus important dans le cas du calendrier à journées convenables que dans celui du calendrier à journées désignées. Toutefois, il est impossible de déterminer si l'utilisation du calendrier à journées convenables donne lieu à une surestimation du temps passé en-dehors de la maison, ou si le calendrier à journées désignées produit une sous-estimation de ce temps, car on ne sait pas quelle est la vraie réponse. Dans Lyberg (1989), on a demandé au répondant de tenir deux journaux. La tenue du premier journal était basée sur un calendrier à journées désignées et l'autre, sur un calendrier à journées convenables. Cependant, le questionnaire du journal à journées convenables a été administré par l'intervieweur, tandis que celui du journal à journées désignées a été auto-administré plusieurs jours après l'interview basée sur les journées convenables. Donc, il est impossible de déterminer si un écart donné est dû à des différences entre les calendriers de prise de contact ou à des effets de mode.

Deux études (Lyberg 1989; Laaksonen et Pääkkönen 1992) portent sur l'effet qu'a la remise de l'interview sur les taux de réponse. Dans les deux cas, les auteurs constatent que la remise de l'interview fait augmenter le taux de réponse. Laaksonen et Pääkkönen (1992) admettent aussi qu'il est difficile de déterminer si le fait de repousser l'interview introduit un biais. Leurs résultats montrent que les répondants qui font repousser l'interview consacrent moins de temps aux tâches ménagères et à l'entretien, et plus de temps au magasinage et aux courses. Cependant, ils ne peuvent établir avec certitude si ces différences sont dues

au biais introduit par la remise de l'interview, à l'hétérogénéité inobservée qui est corrélée à la probabilité de remise de l'interview ou simplement à un bruit aléatoire. Quoi qu'il en soit, ils soutiennent que les différences sont très faibles, si bien que tout biais éventuel est petit.

L'un des avantages du calendrier à journées convenables est qu'il est possible de procéder à de nombreuses tentatives de prise de contact en une brève période. En revanche, le calendrier à journées désignées – tel que proposé – ne permet de procéder qu'à une seule tentative par semaine. Donc, il est naturel de se demander s'il serait déraisonnable de modifier le calendrier à journées désignées pour permettre une certaine forme de substitution du jour de la semaine. Par exemple, si le répondant ne peut être rejoint le mardi pour faire une déclaration concernant le lundi, serait-il acceptable de communique avec lui, disons, le jeudi en lui demandant de faire une déclaration concernant le mercredi? Ce calendrier modifié permettrait de faire un plus grand nombre de tentatives de prises de contact sans devoir prolonger la période de travail sur le terrain.

Comme ce genre de substitution n'est sensée que si les journées de remplacement sont assez semblables aux journées originales, la première étape consiste à déterminer quelles journées, si tant qu'il y en ait, sont similaires à d'autres. Lors de travaux antérieurs, Stewart (2000) nous a montré que, du lundi au jeudi, les journées sont fort semblables les unes aux autres, que les vendredi sont légèrement différents des jours de fin de semaine, et que le samedi et le dimanche sont très différents des jours de semaine et l'un de l'autre. Donc, il serait raisonnable de permettre des substitutions entre les jours de la semaine, au moins du lundi au jeudi.

Biais d'activité et biais de non-contact

Lors du choix d'une stratégie de prise de contact, nous devons tenir compte de deux types de biais, à savoir le biais d'activité et le biais de non-contact. Le biais d'activité survient lorsque la probabilité de joindre et d'interviewer un répondant potentiel durant une journée particulière est corrélée aux activités de ce répondant durant la journée correspondante pour laquelle il doit produire le journal de ses activités. Notons que dans la suite de l'article, l'expression probabilité d'un contact s'entend de la probabilité d'un contact productif (c'est-à-dire qui aboutit à une interview). Afin d'isoler les effets de diverses stratégies de prise de contact, on suppose que les répondants avec lesquels il y a eu contact acceptent systématiquement de participer à l'interview. Le biais de non-contact survient lorsque des différences d'activités font varier la probabilité d'un contact d'une personne à l'autre. Suivent des exemples numériques simples qui illustrent ces biais.

Exemple 1 – Biais d'activité : Supposons que les journées des répondants potentiels se répartissent en deux catégories, c'est-à-dire les journées à prise de contact difficile (journées PCD) et les journées à prise de contact facile (journées PCF). En outre, supposons que les intervieweurs ne prennent jamais contact avec les répondants durant des journées PCD (c'est-à-dire que $P_D = 0$, où P_D est la probabilité d'un contact durant une journée PCD) et qu'ils prennent toujours contact avec les répondants durant des journées PCF (c'est-à-dire que $P_F = 1$, où P_F est la probabilité d'un contact durant une journée PCF). Enfin, supposons que la probabilité qu'une journée soit une journée PCF est de 0,5, si bien qu'en moyenne, la moitié des journées de chaque répondant potentiel sont PCF et l'autre moitié, PCD. Notons que les répondants potentiels sont tous identiques, en ce sens que la probabilité que toute journée soit une journée PCF est de 0,5 pour tous les répondants potentiels. Pour simplifier, nous supposons que les activités durant une journée particulière peuvent être résumées grâce à un « indice d'activité », I_J, où $I_J = 1 - P_J$ $(J = D, F)$. L'indice d'activité représente le temps consacré aux activités qui sont corrélées négativement la probabilité d'un contact. Donc, les journées PCD sont des journées durant lesquelles plus de temps est consacré aux activités qui ont lieu en-dehors de la maison (travail, magasinage, loisirs, *etc.*), tandis que les journées PCF sont des journées durant lesquelles plus de temps est consacré aux activités entreprises à la maison (travaux ménagers, loisirs passifs, *etc.*). L'indice d'activité réelle moyenne pour la population de répondants possibles est égal à 0,5 $(= 0{,}5 \times 1 + 0{,}5 \times 0)$.

Si l'on utilise un calendrier de prise de contact à journées convenables et que le nombre de rappels n'est pas limité, il y a suréchantillonnage des journées PCD. Pour comprendre pourquoi, examinons en détail les deux séries possibles de contacts. Si la première tentative de contact a lieu durant une journée PCF, il y a prise de contact avec le répondant auquel on demande de déclarer les activités de la journée précédente (la journée du journal). Comme la probabilité qu'il s'agisse d'une journée PCF ou PCD est la même, en moyenne, la moitié des journées du journal seront des journées PCD et l'autre moitié, des journées PCF. Par conséquent, l'indice d'activité moyenne pour les journées du journal de ces répondants est égal à 0,5, c'est-à-dire la même valeur que pour la moyenne de la population. Si, par contre, la journée du premier contact est une journée PCD, aucune interview n'a lieu et on rappelle le répondant le jour suivant. Les tentatives de prise de contact se poursuivent chaque jour jusqu'à ce que le répondant potentiel soit rejoint (durant une journée PCF). Pour ce répondant, l'indice d'activité moyenne pour les journées du journal est égal à 1, parce que le répondant est toujours interviewé durant une journée PCF qui suit directement une journée PCD. Donc, si une journée particulière est une journée PCD (c'est-à-dire durant laquelle le répondant potentiel entreprend beaucoup d'activités en-dehors de la maison), il est plus probable que cette journée soit sélectionnée comme journée de référence. Par conséquent, la probabilité d'interviewer le répondant potentiel durant une journée de référence particulière est corrélée aux activités qui ont lieu cette journée-là. Puisque la moitié des premières tentatives de

prise de contact sont faites durant des journées PCD et l'autre moitié, durant des journées PCF, l'indice d'activité moyenne pour l'échantillon final est égal à 0,75 (= 0,5 × 0,5 + 0,5 × 1).

Exemple 2 – Biais de non-contact : Supposons maintenant que les répondants potentiels diffèrent l'un de l'autre en ce qui a trait aux probabilités d'un contact et que, pour chaque personne, cette probabilité ne varie pas d'un jour à l'autre. Supposons aussi que la moitié des répondants potentiels sont rejoints durant des journées PCD, avec la probabilité d'un contact $P_D = 0,25$, et que l'autre moitié sont rejoints durant des journées PCF avec probabilité d'un contact $P_F = 0,75$. Si nous essayons de contacter chaque répondant potentiel quatre fois, étant donné ces probabilités, pratiquement tous les répondants potentiels PCF (99,6 %) seront rejoints. En revanche, 68,4 % seulement des répondants potentiels PCD le seront. Le taux global de contacts sera de 84 % (99,6 % × 0,50 + 68,4 % × 0,50), mais l'échantillon final ne sera pas représentatif, puisque 59,3 % seront des répondants PCF et 40,7 % seulement des répondants PCD. Par conséquent, les estimations basées sur cet échantillon auront tendance à sous-estimer le temps consacré aux activités entreprises par les personnes PCD et à surestimer celui consacré aux activités entreprises par les personnes PCF.

Les biais décrits plus haut ne se produisent pas uniquement dans le cas des enquêtes sur l'emploi du temps. Dans la plupart des enquêtes, des mesures sont prises pour réduire au minimum le biais de non-contact, mais moins d'attention est accordée au biais d'activité. Par exemple, outre l'objectif principal consistant à recueillir des données biographiques sur l'emploi, les enquêtes nationales longitudinales incluent quelques questions sur les activités de la population active (emploi et heures de travail) durant la semaine qui a précédé l'interview. Comme le moment de l'interview est généralement fixé à la meilleure convenance du répondant, les activités de ce dernier durant la semaine de référence seront corrélées à la probabilité de l'interviewer au sujet de cette semaine de référence. Le raisonnement intuitif qui sous-tend cette corrélation est exactement le même que pour l'exemple 1. Cette corrélation introduit un biais dans les estimations du nombre d'heures de travail, mais la direction de ce biais est indéterminée. Le nombre d'heures travaillées par semaine a tendance à être surestimé chez les personnes pour lesquelles on a dû repousser l'interview à cause d'un horaire de travail chargé et à être sous-estimé chez celles qui étaient en congé. Le biais d'activité pose aussi un problème dans le cas des enquêtes sur les voyages. Le temps passé en-dehors de la maison a tendance à être surestimé si l'on demande aux répondants des renseignements au sujet, disons, des quatre semaines qui ont précédé l'interview. On peut éliminer ce biais en demandant aux répondants de donner des renseignements au sujet d'une période de référence fixe.

Soulignons qu'il faut accorder plus d'attention au biais de non-contact dans le cas des enquêtes sur l'emploi du temps que dans celui d'autres enquêtes, car, contrairement à la plupart des autres, celles sur l'emploi du temps ne permettent pas la réponse par procuration. Si l'on pouvait accepter ce genre de réponse, les données sur les personnes PCD pourraient être recueillies par procuration auprès de personnes plus facile à rejoindre. Cette mesures affaiblirait la corrélation entre les activités de la personne et la probabilité de recueillir des renseignements au sujet de cette personne.

Le reste de l'article est présenté comme suit. À la section 2, nous décrivons quatre stratégies de prise de contact et nous utilisons des simulations simples pour évaluer le biais d'activité associé à chacune d'elles. À la section 3, nous étoffons les simulations à l'aide de données provenant du Work Schedule Supplement to the Current Population Survey de mai 1997 et de la Time Diary Study réalisée en 1995 par l'Université du Maryland et nous examinons comment le biais varie selon le genre d'activité. En outre, nous décomposons le biais global, pour évaluer la contribution relative du biais d'activité et du biais de non-contact. Enfin, à la section 4, nous résumons les résultats et faisons des recommandations.

2. Stratégies de prise de contact, activités corrélés et biais d'activité

À la présente section, nous comparons les biais d'activité associés au calendrier de prise de contact à journées convenables et à chacune des trois versions du calendrier à journées désignées. Ces calendriers sont définis comme suit :

1. Journées convenables (JC) : tentative de prise de contact avec les répondants potentiels chaque jour après la première tentative, jusqu'à ce qu'on joigne le répondant potentiel ou que la période de travail sur le terrain se termine.
2. Journées désignées (JD) : tentative unique de prise de contact avec les répondants (pas de tentatives subséquentes).
3. Journées désignées avec remise (JDR) : tentatives de prise de contact avec les répondants potentiels le même jour de la semaine que celui de la première tentative, jusqu'à ce qu'on joigne le répondant potentiel ou que la période de travail sur le terrain se termine (tel que recommandé par Kalton 1985).
4. Journées désignées avec remise et substitution (JDRS) : tentatives de prise de contact avec les répondants potentiels un jour sur deux après la première tentative jusqu'à ce qu'on joigne le répondant potentiel ou que la période de travail sur le terrain se termine.

La stratégie JDRS se fonde sur la supposition d'une alternance de tentatives de prise de contact les mardi/jeudi et les mercredi/vendredi. Le fait que les journées de prise de

contact de la première semaine soient les mardi/jeudi ou mercredi/vendredi dépend de la journée de départ, qui est attribuée au hasard.

Comme l'illustre l'exemple 1, il est facile de montrer que le calendrier à journées convenables peut introduire un biais d'activité dans les estimations de l'emploi du temps si la probabilité de base d'un contact est la même pour chaque journée (0,5), à part le bruit aléatoire (+0,5 avec une probabilité de 1/2 ou –0,5 avec une probabilité de 1/2). Même si Stewart (2000) montre que les journées comprises entre le lundi et le jeudi sont fort semblables en moyenne, il est probable que, pour certaines personnes, les probabilités d'un contact varient systématiquement selon le jour de la semaine. Par exemple, il pourrait être difficile de rejoindre certaines personnes les lundi, mercredi et vendredi de chaque semaine. Cette variation systématique rend considérablement plus difficile le dépistage d'un biais dans les estimations d'échantillon, ainsi que la détermination de la direction et de la grandeur de ce biais. On pourrait modéliser des stratégies de prise de contact et résoudre analytiquement pour le biais sous diverses hypothèses au sujet du profil des probabilités d'un contact. Il s'agirait toutefois d'un processus fastidieux, car chaque hypothèse concernant le profil des probabilités d'un contact selon la journée nécessiterait une solution distincte. Par contre, les simulations informatiques constituent un moyen idéal d'évaluer le biais associé à diverses stratégies de prise de contact sous diverses hypothèses quant au profil des probabilités d'un contact. Le programme informatique est plus simple et produit des résultats plus intuitifs que la solution analytique. En outre, il est très facile de le modifier en vue d'utiliser des profils différents. À la section 3, nous ajoutons une touche de réalisme aux simulations en intégrant des données réelles sur l'emploi du temps – ce qui serait impossible si on adoptait la démarche analytique.

Simulations

La stratégie de simulation est très simple. En premier lieu, nous avons créé, pour chacun des 10 000 répondants potentiels, des « données » pour une période de quatre semaines. Afin de nous concentrer sur les stratégies de prise de contact, nous ne nous sommes pas préoccupés des méthodes d'échantillonnage et avons supposé que l'échantillon de répondants potentiels était représentatif de la population. Les simulations sont conçues pour permettre de comparer les calendriers de prise de contact susmentionnés, de sorte que nous supposons que la « semaine » compte cinq jours. Nous avons limité les journées admissibles pour produire le journal de l'emploi du temps du lundi au jeudi, car, comme nous l'avons mentionné plus haut, ces journées sont celles qui se ressemblent le plus. L'étape suivante consistait à simuler les tentatives de prise de contact avec ces répondants conformément aux quatre calendriers de prise de contact décrits plus haut. Enfin, nous avons comparé les estimations ainsi obtenues aux valeurs d'échantillon.

Pour simplifier les simulations, nous avons produit un résumé analytique d'activités spécifiques, comme dans les exemples qui précèdent, et caractérisé chaque journée à l'aide d'un indice d'activité, I_J, (J = D, F) qui varie de 0 à 1. Cet indice d'activité est donné par $I_J = 1 - P_J$ où P_J représente la probabilité de rejoindre et d'interviewer le répondant. Pour simuler la variation des activités d'une journée à l'autre, nous avons posé que la probabilité d'un contact lors d'une journée particulière est donnée par :

$$P_J = \overline{P}_J + \varepsilon,$$

où $\overline{P}_J$ est la probabilité moyenne d'un contact durant une journée PCD (J = D) ou PCF (J = F), et $\varepsilon \sim U(-\hat{\varepsilon}, \hat{\varepsilon})$. Nous supposons que $\overline{P}_D < \overline{P}_F$, autrement dit, en moyenne, les répondants sont moins susceptibles d'être rejoints durant les journées PCD que durant les journées PCF. Pour nous assurer que les probabilités d'un contact soient comprises dans l'intervalle [0,1], nous établissons $\hat{\varepsilon}$ de sorte que $\hat{\varepsilon} < \min(\overline{P}_D, 1 - \overline{P}_F)$.

Nous pouvons formuler de nombreuses hypothèses concernant le profil des activités selon le jour de la semaine. Le cas le plus simple est celui où toutes les journées sont identiques, sauf en ce qui concerne le bruit aléatoire. Toutefois, comme nous l'avons mentionné plus haut, il se peut qu'il soit systématiquement plus difficile de joindre les répondants potentiels certains jours que d'autres. Pour couvrir une grande gamme de profils d'activités, nous avons effectué les simulations sous les huit hypothèses qui suivent en ce qui concerne la répartition des journées PCD et PCF durant chacune des quatre semaines.

1. Les valeurs réelles de l'indice d'activité sont distribuées comme U(0,1), de sorte que la valeur moyenne est 0,5.
2. Les deux premières journées de chaque semaine sont des journées PCD et les trois dernières, des journées PCF (DDFFF).
3. Les trois premières journées de chaque semaine sont des journées PCD et les deux dernières, des journées PCF (DDDFF).
4. Les quatre premières journées de chaque semaine sont des journées PCD et la dernière est une journée PCF (DDDDF).
5. La première journée de chaque semaine est une journée PCF et les quatre dernières sont des journées PCD (FDDDD).
6. Les deux premières journées de chaque semaine sont des journées PCF et les trois dernières, des journées PCD (FFDDD).
7. Les trois premières journées de chaque semaine sont des journées PCF et les deux dernières, des journées PCD (FFFDD).

8. Pour la moitié de l'échantillon, les lundi, mercredi et vendredi sont des journées PCD et les mardi et jeudi, des journées PCF (DFDFD). Pour l'autre moitié de l'échantillon, la situation est inversée (FDFDF).

Dans le cas du profil 1, la probabilité de base de joindre le répondant est la même dans tous les cas, si bien que la variation des probabilités est due entièrement aux termes aléatoires. Dans le cas des profils 2 à 7, les journées PCD sont regroupées soit au début soit à la fin de la semaine. Dans le cas du profil 8, les probabilités de base alternent entre celles des journées PCD et PCF.

Pour étudier le biais d'activité, nous avons exécuté des simulations distinctes pour chacun des huit profils décrits plus haut. Donc, pour une simulation particulière, tous les individus affichent le même profil de probabilités de base.

Le tableau 1 donne les résultats provenant d'un sous-ensemble représentatif des 153 simulations exécutées. Les quatre premières colonnes montrent la probabilité moyenne d'un contact durant les journées PCD et PCF, la valeur de $\hat{\varepsilon}$ et l'indice d'activité réelle moyenne. Les autres colonnes contiennent des estimations du biais associé aux quatre calendriers de prise de contact. Pour calculer le biais, nous avons soustrait la quantité réelle de temps consacré à chaque activité de la quantité estimée, puis exprimé la différence en pourcentage de la valeur réelle. L'astérisque qui figure à côté de certaines entrées indique que le biais est statistiquement non nul au niveau de signification de 5 %.

Profil 1 – Probabilités de base identiques avec bruit aléatoire

Ce profil est essentiellement le même que celui de l'exemple numérique décrit plus haut. Le résultat principal est que tous les calendriers de prise de contact produisent des estimations non biaisées de l'indice d'activité moyenne

Tableau 1
Biais d'activité associé à chaque stratégie de prise de contact sous diverses hypothèses concernant la corrélation des activités des diverses journées

Profil d'activités	Probabilité moyenne d'un contact		$\hat{\varepsilon}$	Indice d'activité réelle moyenne	Biais estimé (exprimé en pourcentage de l'indice d'activité réelle)			
	Journées à contact difficile	Journées à contact facile			JC	JD	JDR	JDRS
Probabilités de base identiques								
	0,50		0,10	0,500	0,7*	– 0,1	0,0	0,1
	0,50		0,30	0,500	5,3*	– 0,3	0,1	0,2
	0,50		0,50	0,500	15,1*	– 0,9	0,4	0,7
Probabilités de base groupées								
DDFFF	0,75	0,25	0,05	0,500	0,7	– 10,7*	– 4,7*	– 13,8*
	0,75	0,25	0,25	0,500	5,2*	– 10,9*	– 4,8*	– 13,9*
	0,60	0,40	0,05	0,500	– 0,1	– 2,2*	– 0,7*	– 2,8*
	0,60	0,40	0,20	0,500	2,5*	– 2,6*	– 0,7*	– 2,5*
DDDFF	0,75	0,25	0,05	0,625	– 2,7*	– 9,7*	– 4,0*	– 12,7*
	0,75	0,25	0,25	0,625	0,8	– 10,3*	– 4,1*	– 12,8*
	0,60	0,40	0,05	0,550	– 0,4*	– 1,8*	– 0,6*	– 2,5*
	0,60	0,40	0,20	0,550	1,9*	– 2,4*	– 0,5	– 2,2*
DDDDF	0,75	0,25	0,05	0,750	0,1	– 0,1	0,1	0,0
	0,75	0,25	0,25	0,750	2,3*	– 0,5	0,2	0,2
	0,60	0,40	0,05	0,600	0,1*	0,0	0,0	0,0
	0,60	0,40	0,20	0,600	1,9*	– 0,3	0,2	0,2
FDDDD	0,75	0,25	0,05	0,625	1,7*	1,0	1,4*	0,7
	0,75	0,25	0,25	0,625	4,2*	– 0,3	1,2*	0,7
	0,60	0,40	0,05	0,550	1,1*	0,3	0,5*	0,3
	0,60	0,40	0,20	0,550	2,9*	0,0	0,6*	0,4
FFDDD	0,75	0,25	0,05	0,500	– 18,2*	– 17,1*	– 4,3*	– 21,7*
	0,75	0,25	0,25	0,500	– 15,9*	– 17,9*	– 4,5*	– 20,9*
	0,60	0,40	0,05	0,500	– 2,0*	– 2,2*	– 0,4	– 2,6*
	0,60	0,40	0,20	0,500	– 0,4	– 2,4*	– 0,3	– 2,6*
FFFDD	0,75	0,25	0,05	0,375	– 16,6*	– 17,6*	– 5,5*	– 20,3*
	0,75	0,25	0,25	0,375	– 11,4*	– 17,6*	– 5,6*	– 19,6*
	0,60	0,40	0,05	0,450	– 2,0*	– 2,3*	– 0,4	– 2,5*
	0,60	0,40	0,20	0,450	0,0	– 2,5*	– 0,5	– 2,5*
Probabilités de base alternées								
DFDFD/FDFDF	0,75	0,25	0,05	0,500	31,5*	26,4*	9,6*	28,5*
	0,75	0,25	0,25	0,500	34,7*	26,5*	9,7*	29,4*
	0,60	0,40	0,05	0,500	5,6*	4,5*	1,3*	5,1*
	0,60	0,40	0,20	0,500	7,8*	4,3*	1,2*	5,1*

Nota : Les astérisques indiquent que l'indice estimé d'activité moyenne est statistiquement différent de la valeur réelle au niveau de signification de 5 %.

sauf le calendrier à journées convenables (JC). Comme prévu, le calendrier JC donne lieu à une surestimation de l'indice d'activité moyenne. Mais, par-dessus tout, lorsqu'on utilise le calendrier JC, l'indice estimé d'activité moyenne – et donc le biais, lorsque les activités ne sont pas corrélées d'une journée à l'autre – est corrélé positivement à la variance de ε. Puisque la variance augmente de 0,003 ($\hat{\varepsilon} = 0{,}1$) à 0,083 ($\hat{\varepsilon} = 0{,}5$), le biais augmente pour passer de moins de 1 % à 15 %. On peut voir ce que signifie intuitivement ce résultat en notant qu'une réalisation négative importante de ε pour une journée particulière rend moins vraisemblable que le répondant soit rejoint cette journée-là et, donc, plus vraisemblable que cette journée devienne la journée du journal. Aucun autre calendrier de prise de contact n'est sensible à la variance de ε.

Profils 2 à 7 – Probabilités de base groupées

Les résultats sont mixtes lorsque les journées PCD sont groupées soit au début soit à la fin de la semaine. Dans les simulations où la différence $\overline{P}_F - \overline{P}_D$ est assez faible (0,2), tous les calendriers de prise de contact donnent d'assez bons résultas. La valeur absolue du biais est inférieure à 3 % dans tous les cas. Cependant, si $\overline{P}_F - \overline{P}_D$ est assez grande (0,5), les écarts entre les biais associés, respectivement, à chaque calendrier de prise de contact sont significatifs. Le calendrier JDR est celui qui donne les meilleurs résultats dans l'ensemble. Le biais n'excède 5 % (en valeur absolue) que pour le profil 7 (FFFDD), pour lequel il est égal à – 5,5 %. Par contre, si l'on utilise les calendriers JD et JDRS, le biais est de l'ordre de 10 à 14 % pour les profils 2 (DDFFF) et 3 (DDDFF), et de l'ordre de 16 à 20 % pour les profils 6 (FFDDD) et 7 (FFFDD). La différence entre les calendriers JD et JDRS, d'une part, et le calendrier JDR, d'autre part, pour ces profils est significative, statistiquement et en pratique. Pour les profils 4 (DDDDF) et 5 (FDDDD), le calendrier JDR donne des résultats un peu moins bons que les calendriers JD et JDRS, mais le biais est si faible (inférieur à 1,5 %) que la différence n'a aucune signification pratique. Le calendrier JC donne d'un peu meilleurs résultats que les calendriers JD et JDRS. Le biais est inférieur à 5 %, sauf pour les profils 6 et 7, où le biais est de l'ordre de 11 à 18 %. Comme pour le profil 1 susmentionné, l'indice estimé d'activité moyenne augmente avec la variance de ε pour le calendrier JC, mais non les autres. Et, comme le montre le tableau 1, pour les profils où le biais est négatif (profils 6 et 7), une augmentation de la variance de ε réduit le biais.

Profils 8 – Probabilités de base alternées

Tous les calendriers de prise de contact produisent des estimations biaisées, car les journées PCF sont sous-échantillonnées. Comme plus haut, tous les calendriers donnent d'assez bons résultats si la différence $\overline{P}_F - \overline{P}_D$ est assez faible. Le biais est de l'ordre de 5 à 8 % pour tous les calendriers, sauf le calendrier JDR, pour lequel il est d'environ 1 %. Cependant, si $\overline{P}_F - \overline{P}_D$ est importante, tous les calendriers de prise de contact produisent un biais significatif. Pour le calendrier JDR, le biais est plus important que pour les autres profils (environ 10 %), mais il est plus faible que la valeur de 25 à 35 % observée pour les autres calendriers. De nouveau, ces différences sont significatives statistiquement et en pratique.

Le calendrier JDRS produit un biais d'activité important parce que les tentatives de prise de contact ont lieu lors de deux journées PCD, puis lors de deux journées PCF (ou inversement). Ce profil aboutit à des prises de contact avec les répondants durant une fraction assez importante des journées PCF et, donc, les journées du journal comptent un nombre disproportionné de journées PCD. Naturellement, si l'on modifie le calendrier JDRS de sorte que les tentatives de contact aient lieu les deux mêmes journées chaque semaine, le biais est virtuellement nul.

Ces simulations montrent clairement que le biais d'activité associé à chaque stratégie de prise de contact dépend du profil journalier des activités, des probabilités d'un contact durant une journée PCD ou une journée PCF, et de la variance de ces probabilités. Cependant, il est également évident que le calendrier JDR donne de meilleurs résultats que les autres, quel que soit le profil d'activité supposé. Si l'on considère chaque profil comme une catégorie différente de répondants, alors le biais global (qui inclut à la fois le biais d'activité et le biais de non-contact) dépend de la fréquence relative de chaque catégorie dans la population. Des renseignements sur l'incidence de chaque catégorie permettrait de mesurer le biais global et, pour chaque stratégie, de décomposer ce dernier en une fraction due au biais d'activité et une fraction due au biais de non-contact. Nous examinons cette question à la section suivante.

3. Simulations étoffées

Si l'on est prêt à émettre certaines hypothèses supplémentaires, il est possible d'étoffer les simulations à l'aide de données provenant d'autres sources. La première hypothèse est que les horaires de travail des individus constituent une approximation raisonnable des profils d'activité des journées PCD et PCF, de sorte que les journées de travail correspondent aux journées PCD et les journées de congé, aux journées PCF. La deuxième hypothèse est qu'il est possible de reproduire la semaine d'un individu donné en prenant une journée de la semaine de cinq individus distincts.

Nous avons utilisé les données du Work Schedule Supplement de mai 1997 à la Current Population Survey (CPS) pour obtenir des renseignements sur les horaires de travail individuels. Notons que, comme il faut connaître la prévalence de chaque type d'horaire pour la population complète, nous avons également inclus les personnes qui ne travaillent pas. Le tableau 2 donne les profils des journées de travail (W) et de congé (N) basés sur la CPS de mai 1997. Deux de ces profils englobent environ 88 %

d'individus. Quarante-huit pour cent travaillent les cinq jours de semaine et 39 % ne travaillent aucun jour de la semaine. En outre, 4 % travaillent quatre jours par semaine et ont congé le vendredi ou le lundi. Les autres ne présentent pas de profil discernable. Pour simplifier les simulations, nous avons supposé que les individus travaillaient les cinq jours de semaine (travailleurs) ou qu'ils ne travaillaient aucun jour de semaine (non-travailleurs).

Tableau 2
Distribution des horaires de travail

Profil d'activité					Pourcentage	Pourcentage cumulatif
L	Ma	Me	J	V		
–	–	–	–		39,40	39,40
W	W	W	W	W	48,11	87,51
W	W	W	W	–	2,63	90,14
–	W	W	W	W	1,63	91,77
W	W	W	–	–	0,81	92,58
W	W	–	–	–	0,26	92,84
–	–	–	W	W	0,37	93,21
–	–	W	W	W	0,68	93,89
W	–	W	–	W	0,49	94,38
–	W	–	W	–	0,25	94,63
–	–	–	–	W	0,51	95,14
W	–	–	–	–	0,25	95,39
W	W	–	W	W	0,73	96,12
W	–	–		W	0,36	96,48
W	–	–	W	W	0,70	97,18
Autres profils					2,82	100,00
Total					100,00	

Nota : Un « W » indique une journée de travail et un « – » indique une journée de congé. Totalisations de l'auteur d'après le Work Schedule Supplement de mai 1997 à la CPS. Les observations ont été pondérées en se servant des poids du supplément. La taille de l'échantillon est de 89 746 observations.

Pour produire des renseignements pour les activités individuelles, nous nous sommes servis de données provenant de l'EPA Time Diary Study de 1992–1994 réalisé par l'Université du Maryland. Cet ensemble de données contient des journaux sur l'emploi du temps pour un ensemble de 7 408 adultes (voir Triplett 1995). Comme chaque personne n'a été interviewée qu'une seule fois, il n'existe qu'une seule observation par personne. Nous avons utilisé la méthode d'échantillonnage répétée qui suit pour produire des données correspondant à huit semaines pour un échantillon de 18 974 « individus ». Nous avons réparti les journées du journal entre les journées de travail et les journées de congé. Nous avons considéré qu'une journée du journal correspondait à une journée de travail si l'individu avait fait un travail rémunéré durant la journée en question. Les journées de travail ont été attribuées aux travailleurs et les journées de congé ont été attribuées aux non-travailleurs. Les lundi ont été tirés des observations faites le lundi, les mardi, des observations faites le mardi, *etc*. Aucune observation n'a été utilisée plus d'une fois pour une personne particulière, mais la même observation pouvait être utilisée pour plus d'une personne. Les proportions finales d'échantillon paraissent assez semblables aux proportions basées sur la CPS. En tout, 58 % de personnes de l'échantillon final étaient des travailleurs et 42 %, des non-travailleurs, valeurs qui sont raisonnablement proches du rapport entre les travailleurs et les non-travailleurs (1,38 c. 1,23) calculé d'après les données de la CPS.

Pour calculer les probabilités d'un contact, il a fallu émettre une troisième hypothèse. À l'instar de Pothoff, Manton et Woodbury (1993), nous avons supposé que la probabilité d'un contact était égale au nombre de minutes consacrées à une activité effectuée à la maison (sauf dormir) divisé par le temps consacré à toutes les autres activités que le sommeil. Ce processus de calcul de la probabilité d'un contact présente deux propriétés importantes : 1) la probabilité d'un contact pour une journée donnée est liée aux activités réalisées cette journée-là et 2) l'un des groupes de répondants potentiels (travailleurs) a une probabilité plus faible de contact productif (0,36 c. 0,72).

Les tableaux 3a et 3b résument les estimations du biais calculées d'après les simulations étoffées. Le tableau 3a montre les estimations du biais en supposant que la période de travail sur le terrain est de quatre semaines et le tableau 3b montre les mêmes estimations en supposant que cette période est de huit semaines. Dans chaque tableau, les quatre premières colonnes contiennent les estimations du biais associées aux quatre calendriers de prise de contact. Les entrées pour chaque calendrier et chaque activité à un chiffre incluent les estimations du biais d'activité pour les travailleurs et les non-travailleurs, et une estimation du biais global. Comme ce dernier inclut le biais de non-contact, il est possible qu'il soit plus grand (ou plus petit) que le biais d'activité pour l'un ou l'autre groupe. Nous avons calculé le biais de la même façon que pour la simulation précédente. Comme pour l'ensemble précédent de simulations, un astérisque indique que le biais est significativement non nul au niveau de signification de 5 %. La cinquième colonne montre le temps réellement consacré à chaque activité, selon le groupe et globalement.

La comparaison des tableaux 3a et 3b montre que, sauf pour la stratégie de prise de contact lors des journées désignées (JD) pour laquelle la période de travail sur le terrain n'a pas d'importance, la différence principale est que le biais global est plus petit lorsque la période de travail sur le terrain est de huit semaines. Ce biais global plus faible est dû principalement au nombre accru de tentatives de prise de contact qui fait augmenter de façon disproportionnée la probabilité que les travailleurs soient rejoints et rend l'échantillon plus représentatif (voir le tableau 4). En revanche, les estimations du biais d'activité associé aux diverses stratégies de prise de contact ne sont pas sensibles à la longueur de la période durant laquelle sont faites les tentatives de contact. Le reste de la discussion se concentrera sur les résultats du tableau 3b.

Tableau 3a
Biais estimatif – Simulations étoffées (période de travail sur le terrain de quatre semaines)

Activité/Situation d'emploi	JC	JD	JDR	JDRS	Temps consacré à l'activité (réel)
Loisirs passifs					
Non-travailleurs	−8,44*	0,12	−1,54	−1,03	314,72
Travailleurs	−5,40*	1,07	0,43	0,82	152,04
Global	−8,62*	13,56*	2,53*	0,38	220,70
Loisirs actifs					
Non-travailleurs	9,80*	−2,75	0,99	−0,66	65,94
Travailleurs	−0,07	−7,34	−4,69	1,91	26,89
Global	4,03*	11,75*	3,31	1,08	43,37
Divertissement/activités sociales					
Non-travailleurs	19,41*	−2,01	−0,25	−1,20	67,30
Travailleurs	8,63*	7,14	5,21	3,72	27,87
Global	13,11*	15,78*	5,64*	1,37	44,51
Activités organisationnelles					
Non-travailleurs	19,58*	−0,98	9,00	3,84	19,25
Travailleurs	13,77*	6,95	7,17	7,48	8,72
Global	15,24*	15,26*	12,37*	5,99	13,16
Éducation/formation					
Non-travailleurs	32,77*	−0,42	12,54*	8,92*	43,60
Travailleurs	−1,17	7,63	0,57	1,59	13,16
Global	19,17*	22,02*	15,39*	8,00*	26,01
Soins personnels					
Non-travailleurs	−0,50	−0,29	−0,49	−0,44	663,04
Travailleurs	−0,52*	0,01	−0,06	−0,13	580,71
Global	−0,79*	2,20*	0,34	−0,15	615,46
Achats de biens/services					
Non-travailleurs	12,62*	1,35	0,11	−1,28	72,98
Travailleurs	−4,05	4,62	−3,62	−5,43*	23,28
Global	4,67*	22,36*	4,25*	−1,49	44,25
Garde active d'enfants					
Non-travailleurs	−7,89*	5,11	−1,06	−0,54	24,13
Travailleurs	−7,69*	−6,05	−4,09	−0,92	12,64
Global	−9,09*	14,21*	0,77	−0,09	17,49
Travaux ménagers					
Non-travailleurs	−8,88*	1,71	0,33	2,27	169,04
Travailleurs	−10,55*	0,85	−2,03	−0,14	57,92
Global	−11,49*	20,77*	4,53*	2,52*	104,82
Travail rémunéré					
Non-travailleurs	—	—	—	—	—
Travailleurs	2,95*	−0,77	0,25	−0,27	536,77
Global	6,74*	31,44*	−7,74*	−1,87*	310,22

Nota : L'astérisque indique que le biais dans l'estimation du temps consacré à l'activité diffère significativement de zéro au niveau de signification de 5 %.

La stratégie JD ne donne lieu à virtuellement aucun biais d'activité. Pour quelques activités – loisirs actifs, divertissement/activités sociales, activités organisationnelles, éducation/formation et garde active d'enfants pour les travailleurs, et garde active d'enfants pour les non-travailleurs – le biais d'activité est assez important, mais aucune des estimations n'est statistiquement significative. Dans le cas de la stratégie JD, le biais global est assez important pour la plupart des activités, situation qui, comme nous le montrerons plus bas, est due principalement au biais de non-contact.

La comparaison des trois autres stratégies de prise de contact permet de dégager deux tendances. En premier lieu, le biais d'activité est nettement plus faible (et généralement statistiquement non significatif) si l'on utilise la stratégie JDR ou la stratégie JDRS plutôt que la stratégie JC. En deuxième lieu, le biais qui entache les estimations JC suit le profil prévu. Il a tendance à être positif pour les activités qui

ont lieu en-dehors de la maison (loisirs actifs, divertissement/activités sociales, activités organisationnelles, éducation/formation, achat de biens/services et travail non rémunéré) et négatif pour les activités qui ont lieu à la maison (loisirs passifs, soins personnels, garde active d'enfants et travaux ménagers). Ce profil confirme le résultat des études mentionnées dans l'introduction, à savoir que le temps consacré à des activités en-dehors de la maison est plus important dans le cas d'une stratégie de prise de contact durant des journées convenables pour le répondant que dans celui d'une stratégie à prise de contact lors de journées désignées. Mais avant tout, il est maintenant évident que ce résultat est dû à un biais lié à la stratégie de prise de contact durant des journées convenables plutôt qu'à un biais lié aux stratégies de prise de contact durant des journées désignées.

Biais de non-contact

En général, le taux de contacts augmente et l'échantillon devient plus représentatif à mesure que le nombre de tentatives de prise de contact augmente (voir le tableau 4). C'est pour la stratégie JD que le taux de contacts est le plus faible (40 %) et que l'échantillon est le moins représentatif. Dans le cas des calendriers de prise de contact JDR et JDRS, le taux de contacts augmente et l'échantillon devient plus représentatif lorsque la période de travail sur le terrain

Tableau 3b
Biais estimatif – Simulations étoffées (période de travail sur le terrain de huit semaines)

Activité/Situation d'emploi	JC	JD	JDR	JDRS	Temps consacré à l'activité (réel)
Loisirs passifs					
Non-travailleurs	– 8,63*	– 0,09	– 1,62	– 1,21	315,38
Travailleurs	– 5,24*	1,28	0,39	1,10	151,72
Global	– 8,72*	– 13,51*	– 0,35	– 0,31	220,79
Loisirs actifs					
Non-travailleurs	10,62*	– 2,03	1,76	0,06	65,46
Travailleurs	0,00	– 7,29	– 3,50	2,21	26,87
Global	4,49*	12,30*	0,50	0,82	43,16
Divertissement/activités sociales					
Non-travailleurs	19,77*	– 1,72	– 0,15	– 0,91	67,10
Travailleurs	8,09*	6,64	5,52	2,76	28,00
Global	13,06*	15,80*	2,47	0,40	44,50
Activités organisationnelles					
Non-travailleurs	18,92*	– 1,53	8,59	3,25	19,36
Travailleurs	14,03*	7,00	3,18	7,25	8,72
Global	14,89*	14,88*	7,14*	4,76	13,21
Éducation/formation					
Non-travailleurs	33,56*	0,18	12,91*	9,55*	43,34
Travailleurs	– 0,72	8,24	0,77	2,01	13,09
Global	19,73*	22,74*	10,29*	7,32*	25,86
Soins personnels					
Non-travailleurs	– 0,50	– 0,29	– 0,48	– 0,44	663,03
Travailleurs	– 0,55*	0,00	– 0,08	– 0,16	580,81
Global	– 0,82*	2,20*	– 0,17	– 0,29	615,51
Achats de biens/services					
Non-travailleurs	12,64*	1,36	– 0,09	– 1,28	72,97
Travailleurs	– 4,41	4,23	– 3,66	– 5,45*	23,36
Global	4.48*	22.23*	– 0.42	– 2.58	44.30
Garde active d'enfants					
Non-travailleurs	– 7.67*	5.36	– 1.04	– 0.31	24.07
Travailleurs	– 8.02*	– 6.18	– 4.98	– 1.65	12.66
Global	– 9.14*	14.30*	– 2.23	– 0.89	17.48
Travaux ménagers					
Non-travailleurs	– 9.02*	1.55	0.20	2.10	169.30
Travailleurs	– 10.55*	0.80	– 2.15	– 0.20	57.95
Global	– 11.64*	20.63*	0.17	1.34	104.94
Travail rémunéré					
Non-travailleurs	—	—	—	—	—
Travailleurs	2.96*	– 0.78	0.30	– 0.26	536.82
Global	6.86*	– 31.44*	– 0.86	– 0.22	310.25

Nota : L'astérisque indique que le biais dans l'estimation du temps consacré à l'activité diffère significativement de zéro au niveau de signification de 5 %.

passe de quatre à huit semaines. L'utilisation d'un calendrier de prise de contact JDRS durant une période de travail sur le terrain de huit semaines (16 tentatives de prise de contact) produit un taux de contacts de 80 % et un échantillon représentatif. Évidemment, l'échantillon obtenu pour le calendrier JDR avec une période de travail sur le terrain de huit semaines est virtuellement identique à celui obtenu avec un calendrier JDRS et une période de travail sur le terrain de quatre semaines.

Biais d'activité c. biais de non-contact

Pour obtenir une image plus précise de la contribution de chaque type de biais au biais global, nous avons décomposé ce dernier en une fraction due au biais d'activité, un fraction due au biais de non-contact et une fraction due à l'interaction entre les deux biais. Pour l'activité *a* et le groupe *g* (travailleurs ou non-travailleurs), le biais global est donné par :

$$F_g X_{ag} - F_g^* X_{ag}^* = F_g^*(X_{ag} - X_{ag}^*) + X_{ag}^*(F_g - F_g^*) + (F_g - F_g^*)(X_{ag} - X_{ag}^*)$$

$$\text{Activité} + \text{Non-contact} + \text{Intéraction}$$

Tableau 4
Sommaire des taux de contact – Simulations étoffées

Période sur le terrain		JC	JD	JDR	JDRS	Valeur réelle
4 semaines	Taux de contacts	89,68	40,35	71,79	78,39	
	Pourcentage de non-travailleurs	40,08	60,07	46,82	43,14	42,21
	Pourcentage de travailleurs	59,92	39,93	53,18	56,86	57,79
8 semaines	Taux de contacts	89,79	40,35	78,87	80,17	
	Pourcentage de non-travailleurs	40,02	60,07	42,88	42,19	42,21
	Pourcentage de travailleurs	59,98	39,93	57,12	57,81	57,79

où F_g est la fraction de l'échantillon dans le groupe *g*, et X_{ag} est le temps consacré à l'activité *a* par le groupe *g*, et les astérisques indiquent les valeurs réelles. On obtient le biais total pour l'activité *a* par sommation de cette expression sur les travailleurs et les non-travailleurs, c'est-à-dire :

$$\sum_{g=W,\,N}(F_g X_{ag} - F_g^* X_{ag}^*) = \sum_{g=W,\,N} F_g^*(X_{ag} - X_{ag}^*) + \sum_{g=W,\,N} X_{ag}^*(F_g - F_g^*) + \sum_{g=w,\,N}(F_g - F_g^*)(X_{ag} - X_{ag}^*),$$

nous pouvons tirer plusieurs enseignements de ces décompositions (présentées au tableau 5). Premièrement, dans le cas du calendrier JC, le biais global est dû entièrement au biais d'activité. Le grand nombre de tentatives de prise de contact garantit virtuellement l'obtention d'un échantillon représentatif, si bien qu'augmenter la période de travail sur le terrain de quatre à huit semaines n'apporte que peu de changements. Par contre, le biais de non-contact représente la totalité du biais dans le cas du calendrier JD. Dans le cas des calendriers JDR et JDRS, il n'existe virtuellement aucun biais d'activité et le biais de non-contact diminue spectaculairement lorsque la période de travail sur le terrain passe de quatre à huit semaines. Il n'est pas étonnant de constater que pour le calendrier JDR, le biais de non-contact associé à une période de travail sur le terrain de huit semaines est à peu près le même que le biais de non-contact obtenu pour le calendrier JDRS avec travail sur le terrain de quatre semaines. Dans ces simulations, l'échantillon devient entièrement représentatif lorsque la période de travail sur le terrain est suffisamment longue pour permettre 16 tentatives de prise de contact. Enfin, la faible importance des termes d'interaction témoigne du fait que les biais d'activité et de non-contact associés à chaque stratégie de prise de contact sont corrélés négativement.

4. Sommaire et recommandations

Les enquêtes téléphoniques sur l'emploi du temps ont des caractéristiques uniques qui compliquent la collecte des données. Contrairement à la plupart des autres enquêtes, celles sur l'emploi du temps ne peuvent s'appuyer sur des réponses par procuration, de sorte qu'il est plus probable que la probabilité de rejoindre un répondant potentiel soit corrélée aux activités de ce dernier. En outre, comme, dans le cas des enquêtes téléphoniques sur l'emploi du temps, on demande aux répondants de déclarer leur emploi du temps de la journée précédente, il se peut que la probabilité d'interviewer le répondant au sujet d'une journée de référence particulière soit corrélée aux activités ayant eu lieu durant cette journée de référence. Nous montrons d'abord comment ces caractéristiques peuvent donner lieu à un biais de non-contact et à un biais d'activité. Puis, à l'aide de deux ensembles de simulations informatiques, nous montrons que l'importance de ces biais dépend de la stratégie adoptée pour prendre contact avec les répondants potentiels.

Le premier ensemble de simulations montre que l'importance du biais associé à un calendrier particulier de prise de contact dépend du profil des journées à prise de contact facile (PCF) et à prise de contact difficile (PCD). Le calendrier de prise de contact à journées désignées avec remise (JDR) donne de meilleurs résultats que les autres calendriers pour tous les profils d'activité étudiés. Ces

Tableau 5
Décomposition du biais – Simulations étoffées

	Période de travail sur le terrain de quatre semaines				Période de travail sur le terrain de huit semaines			
	Biais total	Biais d'activité	Biais de non-contact	Interaction	Biais total	Biais d'activité	Biais de non-contact	Interaction
Loisirs passifs								
JC	– 8,62	– 7,23	– 1,57	0,18	– 8,72	– 7,29	– 1,62	0,19
JD	13,56	0,50	13,16	– 0,10	13,51	0,46	13,24	– 0,18
JDR	2,53	– 0,75	3,40	– 0,11	– 0,35	– 0,83	0,50	– 0,02
JDRS	0,38	– 0,29	0,69	– 0,02	– 0,31	– 0,30	– 0,01	0,00
Loisirs actifs								
JC	4,03	6,27	– 1,92	– 0,32	4,49	6,80	– 1,96	– 0,35
JD	11,75	– 4,40	16,08	0,06	12,30	– 3,92	15,97	0,26
JDR	3,31	– 1,05	4,15	0,20	0,50	– 0,13	0,60	0,03
JDRS	1,08	0,26	0,84	– 0,02	0,82	0,83	– 0,02	0,00
Divertissement/activités sociales								
JC	13,11	15,51	– 1,89	– 0,51	13,06	15,53	– 1,92	– 0,54
JD	15,78	1,30	15,82	– 1,34	15,80	1,32	15,69	– 1,21
JDR	5,64	1,72	4,08	– 0,17	2,47	1,91	0,59	– 0,02
JDRS	1,37	0,58	0,82	– 0,04	0,40	0,42	– 0,02	0,00
Activités organisationnelles								
JC	15,24	17,36	– 1,70	– 0,42	14,89	17,06	– 1,76	– 0,40
JD	15,26	2,05	14,28	– 1,08	14,88	1,72	14,39	– 1,23
JDR	12,37	8,30	3,69	0,39	7,14	6,53	0,54	0,07
JDRS	5,99	5,24	0,74	0,01	4,76	4,77	– 0,02	0,00
Éducation/formation								
JC	19,17	22,84	– 2,49	– 1,18	19,73	23,53	– 2,56	– 1,24
JD	22,02	1,94	20,90	– 0,82	22,74	2,54	20,90	– 0,69
JDR	15,39	9,04	5,40	0,96	10,29	9,36	0,78	0,14
JDRS	8,00	6,78	1,09	0,13	7,32	7,35	– 0,02	0,00
Soins personnels								
JC	– 0,79	– 0,51	– 0,28	0,00	– 0,82	– 0,53	– 0,29	0,00
JD	2,20	– 0,13	2,39	– 0,06	2,20	– 0,13	2,39	– 0,06
JDR	0,34	– 0,26	0,62	– 0,02	– 0,17	– 0,26	0,09	0,00
JDRS	– 0,15	– 0,27	0,12	0,00	– 0,29	– 0,29	0,00	0,00
Achats de biens/services								
JC	4,67	7,55	– 2,39	– 0,49	4,48	7,44	– 2,45	– 0,51
JD	22,36	2,34	20,06	– 0,04	22,23	2,23	20,00	0,00
JDR	4,25	– 1,02	5,18	0,10	– 0,42	– 1,18	0,75	0,01
JDRS	– 1,49	– 2,54	1,04	0,01	– 2,58	– 2,55	– 0,02	0,00
Garde active d'enfants								
JC	– 9,09	– 7,81	– 1,40	0,11	– 9,14	– 7,82	– 1,43	0,10
JD	14,21	0,45	11,72	2,04	14,30	0,53	11,66	2,12
JDR	0,77	– 2,32	3,03	0,07	– 2,23	– 2,69	0,44	0,01
JDRS	– 0,09	– 0,69	0,61	0,00	– 0,89	– 0,87	– 0,01	0,00
Travaux ménagers								
JC	– 11,49	– 9,42	– 2,26	0,18	– 11,64	– 9,51	– 2,32	0,19
JD	20,77	1,43	18,93	0,41	20,63	1,31	18,95	0,37
JDR	4,53	– 0,43	4,89	0,08	0,17	– 0,55	0,71	0,01
JDRS	2,52	1,50	0,99	0,03	1,34	1,36	– 0,02	0,00
Travail rémunéré								
JC	6,74	2,95	3,69	0,11	6,86	2,96	3,79	0,11
JD	– 31,43	– 0,77	– 30,90	0,24	– 31,44	– 0,78	– 30,90	0,24
JDR	– 7,74	0,25	– 7,98	– 0,02	– 0,86	0,30	– 1,16	0,00
JDRS	– 1,87	– 0,27	– 1,61	0,00	– 0,22	– 0,26	0,03	0,00

simulations montrent aussi que les estimations obtenues en utilisant un calendrier de prise de contact à journées convenables (JC) sont sensibles à la variance de la probabilité individuelle (au niveau de la personne) d'un contact. Les estimations du temps consacré aux activités qui sont positivement corrélées à la probabilité d'un contact (par exemple, activité ayant lieu à la maison) diminuent à mesure qu'augmente la variance. En revanche, les estimations produites par d'autres calendriers de prise de contact ne sont pas sensibles à la variance individuelle de la probabilité d'un contact.

Étant donné les résultats des simulations simples, il est évident que le biais global calculé pour les diverses stratégies de prise de contact dépend de la fréquence relative de chaque profil dans la population. Comme il n'existe pas de données directes sur ces profils, le premier

ensemble de simulations a été étoffé à l'aide de données de la CPS sur les horaires de travail et de données réelles sur l'emploi du temps provenant de l'EPA Time Diary Study de 1992 – 1994. Les résultats des simulations étoffées confirment ceux des simulations simples et montrent comment le biais peut affecter les estimations du temps consacré à des activités spécifiques. Comme prévu, la stratégie de prise de contact JC introduit un biais d'activité systématique dans les estimations de l'emploi du temps. Le temps consacré aux activités effectuées à la maison est sous-estimé, tandis que celui consacré aux activités en-dehors de la maison est surestimé. Dans le cas des échantillons produits par les stratégies JDR et JDRS, il n'existe aucun biais d'activité systématique. Les simulations montrent aussi que l'augmentation du nombre de tentatives de prise de contact réduit le biais de non-contact.

Les résultats indiquent clairement que le choix de la stratégie de prise de contact importe et poussent à faire deux recommandations.

Premièrement, les enquêtes sur l'emploi du temps devraient se baser sur un calendrier de prise de contact durant des journées désignées avec possibilité de remise (JDR). Le calendrier JDR crée un biais d'activité plus faible que les autres calendriers de prise de contact pour toutes les hypothèses de profil d'activités testées. Le calendrier JDRS donne des résultats presque aussi bons pour les profils d'activité les plus courants. Toutefois, puisque les taux de contacts et les coûts du travail sur le terrain dépendent du nombre de tentatives de prise de contact, la stratégie JDRS n'est pas plus rentable que la stratégie JDR. Donc, il n'y a aucune raison de préférer la première à la seconde.

Deuxièmement, les enquêtes sur l'emploi du temps doivent inclure des mesures en vue de minimiser le biais de non-contact. Comme ce dernier est en grande partie fonction du nombre de tentatives de prise de contact, un moyen évident de le réduire au minimum consisterait à augmenter le nombre de tentatives de prise de contact. Nous ne nous étendrons pas davantage sur ce point, car d'autres auteurs ont étudié la question en profondeur. Par exemple, Bauman, Lavradas et Merkle (1993) montrent que l'âge et la situation d'emploi sont liés au nombre de rappels et que des rappels supplémentaires produisent un échantillon plus représentatif et, Botman, Massey et Kalsbeek (1989) proposent une méthode pour déterminer le nombre optimal de rappels. Une autre solution consisterait à essayer d'augmenter la probabilité de rejoindre les répondants potentiels. On pourrait pour cela déterminer quand ils sont susceptibles d'être à la maison et les appeler à ce moment- là, ou leur offrir d'appeler eux-mêmes le jour désigné pour leur interview. Un autre moyen de rendre les répondants potentiels « plus disponibles » consisterait à leur offrir un encouragement monétaire. Une démarche moins coûteuse en vue de réduire au minimum le biais de contact serait d'ajuster les poids d'échantillonnage. Pothoff et coll.(1993) montrent que, lorsque la variable mesurée est corrélée (entre individus) à la probabilité d'un contact, la pondération basée sur le nombre de rappels est pratique et efficace. En dernière analyse, la combinaison correcte de ces démarches dépendra des contraintes imposées au gestionnaire de l'enquête.

Remerciements

L'auteur remercie John Eltinge, Mike Horrigan, Anne Polivka, Jim Spletzer et Clyde Tucker de leurs commentaires et suggestions. Les opinions exprimées ici sont celles de l'auteur et ne reflètent pas forcément celles du Bureau of Labor Statistics.

Bibliographie

Bauman, S.L., Lavradas, P.J. et Merkle, D.M. (1993). The impact of callbacks on survey estimates in an annual RDD survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1070-1075.

Botman, S.L., Massey, J.D. et Kalsbeek, W.D. (1989). Cost-efficiency and the number of allowable callbacks in the National Health Interview Survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 434-439.

Harvey, A. (1993). Guidelines for time diary data collection. *Social Indicators Research*, 30, 197-228.

Harvey, A. (1999). Guidelines for time use data collection and analysis. *Time Use Research in the Social Sciences*, (Éds. W.E. Pentland, A.S. Harvey, P. Lawton et M.A. McColl). New York: Kluwer Academic/Plenum Publishers. 19-45.

Kalton, G. (1985). Sample design issues in time diary studies. *Time, Goods, and Well-Being*, (Éds. F.T. Juster et F.T. Stafford). Ann Arbor: University of Michigan, Institute of Social Research. 333-351.

Kinsley, B., et O'donnell, T. (1983). Marking time: methodology report of the Canadian time use pilot study–1981. *Explorations in Time Use* (vol. 1), Ottawa: Department of Communications, Employment and Immigration.

Laaksonen, S., et Pääkkönen, H. (1992). Some methodological aspects on the use of time budget data. *Housework Time in Bulgaria and Finland*, (Éds. L. Kirjavainen, B. Anachkova, S. Laaksonen, I. Niemi, H. Pääkkönen et Z. Staikov). 86-104.

Lyberg, I. (1989). Sampling, nonresponse, and measurement issues in the 1984-85 Swedish Time Budget Survey. *Proceedings of the Fifth Annual Research Conference*: Department of Commerce, Bureau of the Census. 210-238.

Pothoff, R.F., Manton, K.G. et Woodbury, M.A. (1993). Correcting for nonavailability in surveys by weighting based on number of callbacks. *Journal of the American Statistical Association*, 88, 424, 1197-1207.

Statistics Canada (1999). *Overview of the Time Use of Canadians in 1998*. General Social Survey, Numéro de catalogue 12F0080XIE; Ottawa, Canada.

Stewart, J. (2000). Alternative Indexes for Comparing Activity Profiles. Article présenté à 2000 International Association for Time-Use Research Conference, Belo Horizonte, Brazil.

Triplett, T. (1995). Data Collection Methods for Estimating Exposure to Pollutants Through Human Activity Pattern Data: A National Micro-behavioral Approach. mimeo, Survey Research Center, University of Maryland.