

LES ÉCHANTILLONS SUCCESSIFS DE LESLIE KISH ET L'AMERICAN COMMUNITY SURVEY

Charles H. Alexander ¹

RÉSUMÉ

Leslie Kish a longtemps prôné l'adoption de plans de sondage à * échantillons successifs +, ou échantillons à renouvellement complet, avec panels mensuels non chevauchants pouvant être cumulés sur des périodes de diverses longueurs pour des domaines de taille variable. Dans ces conditions, une même enquête peut être utilisée pour atteindre plusieurs objectifs. La nouvelle enquête américaine sur les communautés (*American Community Survey (ACS)*) du Census Bureau est fondée sur un plan de sondage à échantillons successifs de ce genre. Les données qu'elle fournit servent au calcul de moyennes annuelles pour mesurer les variations au niveau de l'État et à celui de moyennes mobiles sur trois ans ou sur cinq ans pour décrire des domaines de plus en plus petits. L'article décrit l'influence exercée par Kish sur l'élaboration de l'American Community Survey, ainsi que certains problèmes méthodologiques d'ordre pratique qu'il a fallu résoudre pour mettre en œuvre le plan de sondage.

Mots-clés : Échantillons successifs; moyennes pluriannuelles; cumuls asymétriques.

1. INTRODUCTION

Tel que défini plus bas, un * plan de sondage à échantillons successifs + permet de créer une enquête unique suffisamment souple pour poursuivre plusieurs objectifs. Le concept a été élaboré par Leslie Kish dans une série d'articles (y compris Kish 1979a, 1979b, 1981, 1983, 1986, 1990, 1997, 1998 et Kish et Verma, 1983, 1986) dans lesquels il a énoncé les principes du cumul spatial et temporel d'information. Kish a prôné l'adoption de ce genre de plan de sondage pour atteindre des objectifs divers (Kish, 1998), particulièrement dans les pays en voie de développement (Kish, 1979b), mais aussi dans le contexte du Recensement des États-Unis (Kish, 1981). Son utilisation personnelle des échantillons successifs, qu'il qualifiait alors d'* échantillon continu +, remonte au moins à 1958 (Kish, 1961); un projet antérieur (Mooney, 1956) est mentionné dans Kish (1998).

L'American Community Survey (ACS), qui est destinée à remplacer l'enquête à * questionnaire détaillé + réalisée habituellement dans le cadre du recensement, s'appuiera sur une forme de plan de sondage à échantillons successifs. Le présent article décrit comment le concept des échantillons successifs est appliqué dans le cas de l'ACS, en tenant compte des objectifs et des considérations opérationnelles propres à l'enquête. Les décisions concernant le plan de sondage de l'ACS illustrent certaines questions que peut soulever

¹ Charles H. Alexander, U.S. Census Bureau, Suitland, Maryland, USA 20233

l'utilisation d'échantillons successifs en général. Elles illustrent aussi les divers niveaux — conceptuel, personnel et pratique — sur lesquels s'est exercée l'influence de Leslie Kish dans le domaine du développement des enquêtes.

2. ÉCHANTILLONS SUCCESSIFS

Un plan de sondage à k échantillons successifs + consiste à sélectionner conjointement k échantillons probabilistes (panels) non chevauchants qui représentent chacun $1/F$ de la population complète. À chaque période, les membres d'un panel particulier sont interviewés. Après k périodes, tous les membres de l'échantillon auront été interviewés. Selon la précision exigée, un seul panel de $1/F$ peut suffire à la production de bonnes estimations pour la population dans son ensemble et, éventuellement, pour certains grands domaines. Pour les domaines plus petits, ou pour obtenir une plus grande précision pour les grands domaines, on peut cumuler un nombre variable de panels consécutifs, jusqu'à concurrence de k/F de la population. Un plan de sondage à échantillons successifs pour lequel $k=F$ porte le nom de * recensement continu +. Pour un échantillonnage successif mensuel, il est naturel que F soit égal à un multiple de 12 et que les cumuls soient trimestriels, semi-annuels, annuels ou pluriannuels.

Les * domaines + s'entendent des zones géographiques ainsi que des sous-groupes démographiques. Kish (1987, section 2.3) présente un cadre de référence pour le compromis entre les niveaux de détail géographique et démographique, pour un niveau donné de précision. Un élément encore plus fondamental du concept des échantillons successifs est celui du * cumul asymétrique + des données, sur des périodes de longueur variable pour des tailles de domaine diverses (Kish, 1990, 1998). Ce dernier fut élargi plus tard pour donner une vue des similarités fondamentales du calcul de la moyenne spatiale et de la moyenne temporelle (Kish, 1998), ainsi que du calcul de la moyenne sur divers domaines démographiques. La souplesse du plan de sondage à échantillons successifs tient au fait qu'il permet de réaliser divers compromis entre les niveaux de détail spatial, temporel et démographique.

Leslie Kish a laissé à ses collègues la tâche de généraliser ses idées en une * théorie de la combinaison de populations + (Kish, 1999a, 2001). Lors de la séance de communications offertes sur la * combinaison d'enquêtes + qu'il a organisée à l'assemblée annuelle de 1999 de l'Institut international de statistique, Leslie a expliqué aux conférenciers que leurs travaux portaient tous sur des aspects différents d'un même problème, qu'ils en soient conscients ou non. La portée de la tâche s'étend aux diverses formes de cumul de données recueillies au moyen d'échantillons successifs, ainsi qu'aux moyens de combiner les données provenant de divers pays afin de produire des statistiques pour de plus grandes entités, comme l'Union européenne. Kish (2001) soutient que ces problèmes sont fondamentalement les mêmes que celui de la combinaison d'information provenant d'expériences différentes (Cochran, 1937, 1954).

3. LE QUESTIONNAIRE DÉTAILLÉ DU RECENSEMENT ET LES OPTIONS INTERCENSITAIRES

Le * questionnaire détaillé + du recensement décennal est la source principale de données infranationales sur les *caractéristiques* de la population et du logement aux États-Unis. Les estimations du *nombre* de personnes et d'unités de logement sont calculées d'après le * questionnaire abrégé + du recensement auquel répondent tous les ménages. Le questionnaire détaillé, pour lequel le taux global d'échantillonnage est de un sur six, fournit

des estimations précises, détaillées², de diverses caractéristiques démographiques et économiques pour les divers États, les grandes villes, ainsi que les grands comtés ou groupes de comtés. Il permet de produire des estimations utiles, quoique moins précises et moins détaillées, pour des régions encore plus petites, comme les petites villes et les réserves indiennes, ainsi que pour les secteurs de recensement, qui comptent, en moyenne, 4 000 habitants. Pour les unités gouvernementales les plus petites, on utilise des taux d'échantillonnage plus élevés, pouvant atteindre un sur deux pour les zones les plus petites, de sorte que l'on puisse produire des estimations utilisables pour ces domaines. Afin de compenser le taux d'échantillonnage plus élevé utilisé pour ces régions, celui appliqué aux secteurs de recensement les plus grands est de un sur huit.

Entre les recensements, les programmes statistiques du gouvernement fédéral fournissent assez peu de renseignements sur les caractéristiques de la population à un niveau infranational. Les dénombremens de base du recensement sont mis à jour grâce à un programme intercensitaire d'estimations démographiques, mais les données sur d'autres caractéristiques démographiques et économiques proviennent principalement d'enquêtes nationales. Dans le cas de la *Current Population Survey* (CPS), qui est l'enquête américaine mensuelle sur la population active, le taux d'échantillonnage est d'environ 1 sur 1 000 et le chevauchement entre les unités d'échantillonnage est considérable d'un mois à l'autre, de sorte que l'échantillon ne peut être cumulé profitablement au cours du temps comme cela est possible dans le cas d'échantillons successifs avec renouvellement complet. Un supplément à la CPS est réalisé une fois par an en mars afin de recueillir des renseignements supplémentaires pour produire des estimations du revenu et de la pauvreté au niveau de l'État, mais la précision et le niveau de détail démographique de ces estimations sont limités. Certains programmes s'appuient sur des méthodes de modélisation fondées sur des dossiers administratifs pour produire des estimations du chômage, ainsi que du revenu et de la pauvreté pour les petites régions, mais ne permettent pas de produire des estimations pour diverses autres caractéristiques.

La nécessité de recueillir plus fréquemment des renseignements pour les petits domaines (ou * communautés +) est admise depuis longtemps (Hauser, 1941; Eckler, 1972, page 212; Bounpane, 1986). Leslie a reconnu à son ami, Philip Hauser, l'honneur d'avoir proposé un * recensement annuel par sondage + en 1941. Kish (1981) a proposé un plan de sondage à échantillons successifs comme moyen d'atteindre cet objectif et a présenté plusieurs options, y compris un plan de sondage à échantillons successifs pour la CPS. Au lieu de cette proposition, un recensement de milieu de décennie a été approuvé pour 1985, mais n'a jamais été financé. Une proposition visant à doubler la taille de l'échantillon de la CPS ne l'a pas été non plus (Tupek et coll. 1990).

L'intérêt du Census Bureau pour des données intercensitaires sur les caractéristiques de la population s'est ravivé à la suite d'une proposition de * programme décennal de recensement + présentée par Herriot et coll. (1989). Ce programme aurait permis de recueillir des données pour différents États pour des années différentes; en dernière analyse, la proposition n'a pas été acceptée. Cependant, le plaidoyer dynamique et éloquent de Roger Herriot quant à l'importance et à la valeur éventuelles des données intercensitaires de niveau infranational a sensibilisé les organismes statistiques fédéraux à la possibilité d'adopter un * nouveau paradigme + pour le cycle décennal de collecte des données. L'échantillonnage successif avec renouvellement complet proposé par Kish a été un facteur définitivement déterminant durant cette période où le Bureau a envisagé de nouvelles approches pour le Recensement de 2000 (voir Bounpane, 1986).

Le Congrès ayant manifesté un regain d'intérêt pour des données intercensitaires sur les caractéristiques de la population (Melnick, 1990; Sawyer, 1993), l'option d'une * mesure continue + pour remplacer le questionnaire détaillé du recensement a été envisagée durant les travaux de recherche pour le Recensement de 2000 qui ont débuté en 1992. Le plan de sondage à échantillons successifs de Kish a finalement été proposé parce qu'il offrait non seulement une certaine souplesse pour la production d'estimations, mais aussi la

² Le terme * précises + a trait aux erreurs d'échantillonnage et le terme * détaillées + signifie que les estimations sont données pour un grand nombre de domaines démographiques dans un domaine géographique.

possibilité de recueillir efficacement des données (Alexander, 1993, 1997; National Academy of Sciences, 1994, 1995). De mémoire, les articles les plus influents ont été ceux de Kish (1981, 1990); ceux de Kish et Verma (1983, 1986) ont également été consultés. La * mesure continue +a été renommée plus tard * *American Community Survey (ACS)* +.

L'ACS proposée n'a pas été adoptée pour le Recensement de 2000, mais, après des essais limités réalisés de 1996 à 1998, la méthodologie de l'ACS a été mise en œuvre dans 36 comtés pour la période de 1999 à 2001, de sorte que l'on puisse comparer les résultats de cette enquête aux données du questionnaire détaillé du Recensement de 2000. Un essai à grande échelle a également été réalisé en 2000, au moyen d'un échantillon

³ Kish (1981) propose comme méthode de rechange celle des * listes successives cumulatives +, mais cette option rendrait onéreux le calcul d'estimations régulières pour tous les domaines les plus petits, comme les secteurs de recensement.

locaux et d'autres sources administratives et grâce au ciblage de nouvelles adresses, dans des régions plus rurales, par des intervieweurs préposés à l'ACS et à d'autres enquêtes. En fait, pour produire les échantillons mensuels, on sélectionne un échantillon annuel d'après le MAF du mois de septembre antérieur et on le divise en 12 panels mensuels. En février, on sélectionne d'après le DSF un échantillon supplémentaire de nouvelles unités que l'on répartit entre les mois restants de l'année.

Le remplacement du questionnaire détaillé du Recensement de 2010 par l'ACS est l'une des composantes du programme de remaniement du Recensement de 2010. Ce programme comprend aussi la modernisation du MAF/TIGER, ainsi qu'un programme de recherche et d'essai précoces en vue d'automatiser, de simplifier et d'améliorer les opérations du Recensement de 2010. En principe, le coût budgété de cette combinaison d'améliorations pour le cycle complet de 10 ans sera inférieur à celui de la répétition des méthodes du Recensement de 2000 en 2010. Ce plan diffère considérablement de la vision de l'ACS décrite dans National Academy of Sciences (1994, chapitre 6; 1995, chapitre 6), où je formulais l'espoir qu'en soi, l'élimination du questionnaire détaillé, sans autres améliorations fondamentales, permette une économie suffisante pour financer l'ACS.

5. CERTAINES VARIANTES DU PLAN DE SONDAGE DE BASE ET CERTAINS PROBLÈMES

5.1 Échantillons en grappes à plusieurs degrés

L'ACS est réalisée auprès d'un échantillon systématique à un degré non mis en grappe, parce que les objectifs comprennent la production annuelle de données pour tous les petits domaines géographiques, comme les secteurs de recensement ou les groupes d'îlots. Les discussions publiées par Kish (1981, 1998) montrent clairement que la méthode des échantillons successifs s'applique aussi à l'échantillonnage par grappes et à l'échantillonnage à plusieurs degrés, ainsi qu'à diverses probabilités de sélection. Cependant, pour que l'on puisse parler d'* échantillons successifs +, il faut que les unités primaires d'échantillonnage forment elles-mêmes des échantillons successifs. Un plan de sondage comportant un ensemble fixe d'unités primaires d'échantillonnage (UPE), avec échantillonnage successif dans chaque UPE, produit un * échantillon cumulatif représentatif +(Kish, 1998).

Leslie insistait sur le fait que la proposition d'Herriot et coll. (1989) ne correspondait pas à ce qu'il entendait par * échantillons successifs +. Pourtant, la proposition semble correspondre à la définition si l'on considère les États comme des UPE. Selon moi, ce fait prouve qu'il est implicitement nécessaire qu'un plan de sondage à échantillons successifs produise un échantillon probabiliste représentatif utile lors de chaque période de référence, pour chaque domaine étudié, condition qui n'est pas vérifiée si les UPE sont des États. Autrement dit, les grappes, ou UPE, doivent être beaucoup plus petites que le plus petit domaine étudié. (Voir Kish, 1998, page 38.)

5.2 Taux d'échantillonnage variable

Selon Kish (1998, section 4), un plan de sondage à échantillons successifs peut prévoir l'application d'un taux d'échantillonnage variable selon la strate. Ce genre de plan peut être compliqué, particulièrement si les taux d'échantillonnage varient avec le temps, car la probabilité conditionnelle de sélectionner une unité (sans remise) pour le $j^{\text{ème}}$ panel dans la $h^{\text{ème}}$ strate dépend des taux d'échantillonnage utilisés pour les panels antérieurs dans cette strate. La situation est encore plus compliquée si la strate varie au cours du temps, par exemple si les limites des unités gouvernementales changent.

Par souci de simplification, dans le cas de l'ACS, nous tirons l'échantillon en deux étapes. À la première, nous sélectionnons des * superéchantillons successifs + en appliquant pour chaque panel et chaque année un même

taux d'échantillonnage, égal au taux le plus élevé nécessaire quelle que soit la strate. À la deuxième étape, l'échantillon initial est sous-échantillonné pour obtenir le taux d'échantillonnage souhaité pour chaque strate pour l'année de référence. La sélection de sous-échantillons, qui permet d'éviter le chevauchement avec la totalité des superéchantillons antérieurs, oblige uniquement à suivre le taux d'échantillonnage pour la première étape.

5.3 Mises à jour de la base de sondage

En pratique, la population varie légèrement pour chaque panel. De nouvelles adresses sont ajoutées à la base de sondage. Certaines anciennes adresses n'existent plus et peuvent être supprimées de la liste, ou y être maintenues et supprimées uniquement après que l'on ait essayé de communiquer avec les personnes vivant à ces adresses. Ceci ne présente aucun problème conceptuel fondamental et signifie qu'un * recensement continu + ne consiste pas nécessairement à communiquer avec toute unité de population ayant déjà existé, puisque certaines unités peuvent disparaître trop rapidement après leur création pour tomber dans l'échantillon.

Pour éviter l'enregistrement des divers taux d'échantillonnage conditionnels appliqués pour différentes * cohortes + d'adresses ajoutées au moment des mises à jour du fichier maître d'adresses exécutées à diverses périodes, nous avons jugé pratique d'attribuer des * échantillons rétrospectifs + artificiels en sélectionnant des adresses à partir de chaque ensemble de nouvelles adresses, non seulement pour le panel courant, mais aussi pour les anciens panels. Ces unités ne sont pas interviewées, puisque la période d'existence du panel auquel elles sont assignées est révolue, mais elles peuvent ainsi être évitées durant la sélection sans remise des futurs panels.

5.4 Que se passe-t-il après le panel k?

Une question qui, autant que je le sache, n'a pas été abordée explicitement par Leslie est la façon de tirer l'échantillon pour le panel k+1. Je crois qu'il supposait que le panel k + 1 serait le même que le panel 1, que le panel k + 2 serait une répétition du panel 2, et ainsi de suite. Cette démarche donne de bons résultats pour un échantillon aléatoire simple, mais ne fonctionne pas aussi bien pour un échantillon systématique destiné à étaler l'échantillon sur une liste triée géographiquement, car, comme la base de sondage évolue avec le temps, le panel 1 ne garde pas son espacement uniforme.

Nous prévoyons sélectionner le panel k + 1 et les panels suivants sous forme d'échantillons systématiques frais. Chacun sera sélectionné de sorte qu'il ne chevauche aucun des k - 1 panels antérieurs, si bien que nous aurons systématiquement k panels consécutifs non chevauchants, sans devoir nous inquiéter du chevauchement avec les panels sélectionnés avant cela.

5.5 Date de référence des questionnaires, étant donné une période étendue d'interviews

Pour chaque panel mensuel de l'ACS, les interviews ont lieu sur une période de trois mois, deux mois étant accordés pour le retour du questionnaire dûment rempli par la poste et le suivi téléphonique, avant le lancement des visites sur place plus coûteuses le troisième mois. Donc, les données recueillies pour juin, par exemple, englobent les réponses rapides par la poste des membres du panel de juin, les réponses tardives par la poste et les interviews téléphoniques pour le panel de mai, ainsi que les cas de suivi sur place pour le panel d'avril. Cette situation oblige à se demander s'il faut que les répondants fournissent les réponses qui correspondent à la date d'envoi par la poste, c'est-à-dire la meilleure option en ce qui concerne le biais d'échantillonnage, ou celles qui correspondent au moment où les questions sont posées, c'est-à-dire la meilleure option en ce qui concerne l'erreur de réponse et d'autres erreurs non dues à l'échantillonnage, particulièrement pour les personnes qui ont changé d'adresse.

Compte tenu de ces compromis concernant la qualité des données, nous avons choisi d'utiliser comme date de référence la * date courante + et de recueillir les données sur les caractéristiques des membres du ménage

au moment de l'interview. Nous avons pris cette décision notamment parce que nous estimons que les erreurs non dues à l'échantillonnage seront plus difficiles à évaluer que le biais d'échantillonnage. En outre, les biais d'échantillonnage qui entachent les estimations mensuelles auront tendance à se compenser au cours de l'année. Il s'agit de l'une des raisons de limiter l'ACS à la production d'estimations annuelles et pluriannuelles.

5.6 Utilisation d'estimations intercensitaires de population pour le contrôle des poids de sondage

Le Census Bureau a mis en place un programme d'estimations démographiques ⁴ intercensitaires fondé sur des modèles démographiques. Ces modèles permettent de mettre à jour les données du recensement antérieur, d'après les données des dossiers de l'état civil et d'autres dossiers administratifs. Ces estimations servent de contrôles indépendants de la pondération, ou de ^{*} facteurs de stratification a posteriori ⁺, pour la plupart des enquêtes-ménages nationales (voir Kish, 1965, pp. 90-92). L'ajustement des poids de sondage afin qu'ils concordent avec les valeurs de contrôle permet de réduire la variance des estimations d'enquête, de rajuster les données pour tenir compte des différences de couverture selon l'âge, le sexe, la race ou l'origine hispanique, et d'augmenter la cohérence d'une enquête à l'autre. De façon analogue, dans le cas du questionnaire détaillé du recensement, on utilise les dénombrements au recensement comme contrôles de la pondération.

Habituellement, aucun contrôle de la pondération n'existait pour les domaines géographiques les plus petits, du moins pas au niveau de détail démographique existant pour les régions plus grandes. La production prévue de totaux de contrôle plus détaillés pour la pondération des données de l'ACS est décrite dans Alexander et Wetrogan (2000). Certaines améliorations résulteront de l'utilisation de meilleures sources de données administratives, mais l'ACS proprement dite fournira des données sur les variations de population qui pourront être intégrées dans les modèles démographiques. La différence entre la ^{*} règle du résident courant + appliquée pour l'ACS et la ^{*} règle du résident usuel + appliquée pour le recensement complique le problème; l'ACS comprend une question au sujet des résidents ayant occupé le logement une partie de l'année afin de permettre la correction pour cette différence de concept. Afin de faciliter cette intégration des données d'enquête dans les modèles démographiques, particulièrement en vue d'élaborer des mesures de l'erreur pour les estimations résultantes, le Census Bureau essaie de mettre au point des versions "statistiques" des modèles démographiques utilisés pour produire les estimations intercensitaires de population. Les efforts en vue d'intégrer les approches statistique et démographique s'inspirent de Purcell et Kish (1979).

6. CUMULS DIFFÉRENTS SELON L'OBJECTIF

Afin d'atteindre l'objectif principal de l'ACS, c'est-à-dire remplacer le questionnaire détaillé du recensement en tant que source de statistiques descriptives détaillées, nous prévoyons cumuler les données de l'ACS sur cinq ans pour obtenir un produit de données comparable au ^{*} fichier sommaire +traditionnellement fondé sur le questionnaire détaillé. Cette période est la plus courte pour laquelle l'erreur d'échantillonnage de l'ACS est jugée raisonnablement proche de celle relative au questionnaire détaillé du recensement. Le fichier de données sur cinq ans engloberait toutes les tailles et toutes les catégories de zones géographiques. En ce qui concerne l'affectation des fonds gouvernementaux basée sur l'évaluation des besoins courants, les études en simulation donnent à penser que le cumul des données sur trois ans serait préférable au cumul sur cinq ans, ce qui revient à sacrifier la précision pour une plus grande actualité (Alexander, 1998).

⁴ Leslie qualifierait ces estimations de ^{*} postcensitaires ⁺, réservant le terme ^{*} intercensitaires ⁺ pour les estimations réalisées entre deux recensements achevés.

Pour les régions individuelles, les données les plus importantes publiées seront les moyennes sur un an pour les régions comptant plus de 65 000 habitants et les moyennes sur trois ans pour celles comptant plus de 20 000 habitants, en plus des moyennes sur cinq ans produites pour toutes les régions. Les estimations moyennes annuelles pour les régions dont la population est inférieure à ces seuils seront offertes pour des utilisations plus complexes dans des modèles de séries chronologiques et pour indiquer les variations importantes des moyennes pluriannuelles, mais nous ne leur accorderons pas autant d'importance dans nos publications ou nos sites Web.

Ces produits de données de l'ACS dont nous prévoyons la publication visent à inciter les analystes à utiliser la même période de cumul lorsqu'ils comparent des régions de tailles différentes, car ne pas le faire serait perçu comme injuste pour les régions plus petites. Ce faisant, nous avons accepté la notion de cumul asymétrique en ce qui concerne les niveaux de détail géographique, mais pas nécessairement à l'intérieur d'un niveau de détail géographique particulier. Par exemple, nous pourrions choisir une période d'une année pour comparer les États, mais nous recommanderions une période de cinq années pour tous les comtés repris dans un tableau comparant les grands et les petits comtés. Cette dernière recommandation nous écarte quelque peu de la démarche de Kish (1998, pp. 42-43) qui permettrait d'utiliser des tableaux contenant des estimations sur un an pour les grands comtés, des moyennes sur trois ans pour ceux de taille moyenne, et des moyennes sur cinq ans pour ceux de petite taille. Il sera intéressant de voir quelle pratique les utilisateurs des données adopteront à cet égard.

7. PONDÉRATION DES ANNÉES LORS DU CUMUL PLURIANNUEL

Kish (1998) fait remarquer qu'il existe plusieurs options pour la pondération des cumuls pluriannuels. S'il existe 10 moyennes annuelles \bar{y}_i , alors il existe de nombreux choix de $\bar{y} = \sum_j w_j \bar{y}_j$, où $\sum_j w_j = 1$, pour les cumuls sur 10 ans.

Pour les cumuls de données de l'ACS sur cinq ans et d'autres cumuls pluriannuels, discutés à la section 6, nous prévoyons accorder un poids égal à toutes les années de référence dans les produits de données publiés types, c'est-à-dire $w_i = 0,2$ pour la moyenne sur cinq ans. Il s'agit là d'un sujet de désaccord avec Kish (1998) qui nous priait de considérer d'autres solutions, particulièrement des poids de la forme $w_{i+1} = Cw_i$, avec $C > 1$.

Lorsque l'on pense à des poids inégaux, l'une des questions sous-jacentes est de savoir quel problème statistique nous essayons de résoudre. Si nous prenons le cumul des données de 2003 à 2007 comme exemple, l'objectif est-il de fournir :

- une estimation directe fondée sur le plan de sondage et de la moyenne historique pour la période de 2003 à 2007;
- une estimation basée sur un modèle pour la valeur de 2007;
- une estimation directe, fondée sur le plan de sondage et de la moyenne historique pondérée pour la période allant de 2003 à 2007, en accordant plus de poids aux années récentes?

Interpréter l'estimation pour la période de 2003 à 2007 comme étant une estimation pour 2007 nécessite un modèle ou des hypothèses au sujet de la série chronologique pour la région en question. Le problème peut être considéré comme étant la combinaison d'une estimation directe pour 2007 avec une prévision pour 2007 fondée sur les années 2003, ..., 2006 à condition que la même formule soit utilisée pour toutes les régions et toutes les caractéristiques afin de maintenir l'additivité dans les tableaux et la comparabilité entre tableaux.

Précédemment, j'ai considéré la décision comme étant un choix entre les deux premiers objectifs et j'ai reculé à l'idée d'adopter la deuxième approche pour l'ACS, en dernière analyse à cause des réserves exprimées par Hansen et coll. (1983, sections 3 et 5.5) concernant l'utilisation d'estimations fondées sur un modèle pour la production de statistiques officielles et d'usage général. Étant donné la variété de statistiques et de zones

géographiques couvertes par l'ACS, il existera inévitablement des cas où le modèle de compromis donne de fort mauvais résultats; l'utilisateur des données pourrait ne pas se rendre compte de cet échec ou en être pleinement conscient. Dans quelle mesure la moyenne de compromis peut-elle être considérée comme une estimation valide pour 2007 lorsque le modèle de compromis donne manifestement de mauvais résultats et quelle mesure de l'erreur convient-il de lui associer? Vu sous cet angle, nous avons recommandé d'utiliser les moyennes pluriannuelles non pondérées comme produits de données types polyvalents et de réserver la série chronologique d'estimations annuelles à l'utilisation dans des modèles chronologiques pour des applications spéciales et à l'interprétation des moyennes pluriannuelles lorsque l'on observe des variations durant la période de cinq ans.

Cependant, après avoir relu l'article de Kish (1998), j'interprète maintenant sa vision de la moyenne pondérée comme correspondant à la troisième formulation, c'est-à-dire un estimateur fondé sur le plan de sondage d'un paramètre de population plus à jour. Cette solution permet d'éviter les questions quant à la validité du modèle aux applications générales. Une question demeure : comment justifier et réaliser une solution consensuelle?. En outre, l'application de poids inégaux a tendance à augmenter les erreurs-types des moyennes pluriannuelles. Toutefois, Kish (1998, p. 40) aura le dernier mot à ce sujet :

**D'importantes questions restent à débattre et à étudier. Peut-être à jamais, le sujet étant devenu une *branche d'activité en croissance. +[Traduction]*

8. NE PAS COMBINER LA CPS ET L'ACS

Leslie a souvent déclaré qu'il était heureux de voir son idée mise en œuvre dans l'ACS, mais je pense qu'il était déçu que nous n'ayons pas essayé de remplacer le questionnaire détaillé du recensement et la CPS par une seule enquête. Contrairement à d'autres questions sur lesquelles nous avons des discussions animées, Leslie avait adopté sur ce point une attitude de *non-intervention +. À mon avis, il considérait cette décision comme une question de compromis concernant la qualité que les organismes gouvernementaux devaient résoudre eux-mêmes. Les deux raisons principales de notre décision sont les suivantes.

Nous ne pouvons évaluer adéquatement le taux mensuel de chômage au moyen d'une enquête postale. La mesure correcte du taux de chômage nécessite des questions complexes que l'on ne pourrait poser au moyen d'un instrument envoyé par la poste, comme celles visant à s'assurer qu'une personne qui *recherche du travail +a activement entrepris des démarches pour trouver un emploi. (Voir Butani et coll. 1999). L'enquête supplémentaire du Recensement de 2000, qui empruntait la méthodologie de l'ACS, a produit une sur-estimation importante du taux national de chômage en 2000 (5,3 % contre 4,0 % pour la CPS). Un écart semblable a été observé pour le Recensement de 1990.

Comparativement à la CPS, les taux mensuels calculés d'après les données des enquêtes postales seraient produits avec un retard important. En outre, le fait qu'il soit impossible de réaliser toutes les interviews par la poste pour un panel donné dans le mois désigné biaise les estimations mensuelles (voir la section 5.5 plus haut). La production de moyennes mobiles trimestrielles au lieu d'estimations mensuelles, solution fréquemment suggérée par Leslie (par exemple, Kish 1999), réduirait ces problèmes, mais le rapport mensuel sur le chômage est un indicateur économique indispensable aux États-Unis.

Le remplacement du questionnaire détaillé du recensement par un autre instrument qu'un questionnaire envoyé par la poste est trop coûteux. Une enquête à échantillons successifs réalisée sur place auprès d'un échantillon suffisamment grand pour remplacer le questionnaire détaillé du recensement devrait être trois à quatre fois plus importante que la CPS. La grandeur de l'échantillon dépend de la taille de la population américaine et du nombre de secteurs de dénombrement pour lesquels des estimations doivent être produites d'après le questionnaire détaillé. Le coût par cas d'une telle enquête serait beaucoup plus élevé que pour la CPS, parce qu'il ne serait pas possible d'utiliser un échantillon en grappes ni de recourir à l'interview téléphonique pour les interviews répétées de mêmes ménages, comme cela est le cas pour la CPS. Le coût total

d'une enquête de ce genre serait plusieurs fois plus élevé que le coût cumulatif de l'ACS proposée et de la CPS.

Parce qu'elle est conçue si rigoureusement pour remplacer un questionnaire détaillé, l'ACS n'illustre pas pleinement la souplesse envisagée par Leslie pour une enquête à échantillons successifs. Dans d'autres circonstances, pour une population plus petite et un besoin moins grand de l'* enquête à questionnaire détaillé + pour produire des estimations pour des domaines très petits, ou dans le cas d'exigences moins rigides quant aux délais de production et aux questions à poser, il sera peut-être possible qu'une enquête sur la population active à échantillons successifs permette de répondre à la demande de données régionales. Grâce à l'ajout d'un panel double ou d'autres composantes (Kish, 1998, pages 40 et 41), on pourrait atteindre une gamme encore plus grande d'objectifs.

9. CONTRIBUTION : CONCEPTUELLE, PERSONNELLE ET PRATIQUE

La longue liste d'articles publiés par Leslie Kish au sujet des échantillons successifs démontre clairement l'intensité et la ténacité de sa campagne en faveur de ce qu'il considérait comme une idée importante. L'évolution de cette idée au fil de ses articles illustre aussi la grande attention qu'il accordait aux questions * conceptuelles + concernant les objectifs fondamentaux de qualité d'une enquête. Qu'essayons-nous de faire? Quel est le rapport entre le choix du plan d'enquête et ce que nous essayons de réaliser, et pourquoi? Cette forme d'orientation, qui est essentielle lors du lancement d'un programme d'enquête quand il faut résoudre les * grandes questions +, permet de faire la distinction entre les idées qui sont abandonnées rapidement et celles qui font leur chemin.

L'appui personnel que Leslie offrait aux autres statisticiens dépassait de loin le cadre de ses articles. Alors que je n'étais nullement l'un de ses collègues les plus proches, il me prodiguait régulièrement des conseils personnels ou des encouragements lorsqu'il le jugeait nécessaire. L'expression * *still rolling* + du titre de la version anglaise de cet article était celui que je donnais aux messages électroniques que je lui envoyais pour le tenir au courant de la situation périlleuse de l'ACS aux diverses étapes du cycle budgétaire annuel, c'est-à-dire la plupart du temps. Il répondait brièvement par courriel, mais transmettait toujours les messages importants sous forme de lettres manuscrites.

Enfin, si l'on s'en tient à ses articles, il est évident que Leslie a toujours été une personne pratique, même dans ses moments de réflexion les plus théoriques, et que ses articles ne peuvent être appréciés comme il se doit si l'on ignore ce qui se passait dans le monde des enquêtes à l'époque où il les a rédigés. Lorsque je réfléchis à ses articles sur les échantillons successifs, je décèle de nombreux commentaires, au sujet de détails aussi bien que de principes généraux, qui visaient à éclairer certaines décisions que le Census Bureau devait prendre à l'époque. J'imagine que l'ensemble de ses travaux contient divers messages destinés spécifiquement à aider quelqu'un, quelque part dans le monde, qui devait prendre à l'époque une décision pratique concernant un plan de sondage.

BIBLIOGRAPHIE

Alexander, C. H. (1993). "A Continuous Measurement Alternative for the U.S. Census", *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 486-491.

Alexander, C. H. (1997), "Questions relatives au plan de sondage de l'American Community Survey et résultats préliminaires des tests", *Recueil Symposium 97 de Statistique Canada: Nouvelles orientations pour les enquêtes et les recensements*, Ottawa, Canada, pp. 211-217.

- Alexander, C. H. (1998), "Recent Developments in the American Community Survey", *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 92-100.
- Alexander, C. H. et Wetrogan, S. (2000). "Integrating the American Community Survey and the Intercensal Demographic Estimates Program." *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 295-300.
- Bounpane, P. (1986), "How Increased Automation Will Improve the 1990 Census", *Journal of Official Statistics*, 4, pp. 545-553.
- Butani, S., Alexander, C., et Esposito, J. (1999). "Using the American Community Survey to Enhance the Current Population Survey: Opportunities and Issues." *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, Statistical Policy Working Paper 29, 3, pp. 102-111.
- Cochran, W.G. (1937). "Problems Arising in the Analysis of a Series of Similar Experiments," *Journal of the Royal Statistical Society (B)*, 4, pp. 102-118.
- Cochran, W. G. (1954). "The Combination of Estimates from Different Experiments," *Biometrics*, 10, pp. 101-129.
- Eckler, A. R. (1972), *The Bureau of the Census*. New York: Praeger Publishers.
- Hansen, M. H., Madow, W. G., Tepping, B. J. (1983). "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys". *Journal of the American Statistical Association*, 78, 384, pp. 776-793.
- Hauser, P. M. (1942), "Proposed Annual Census of the Population", *Journal of the American Statistical Association*, 37, pp. 81-88.
- Herriot, R. A., Bateman, D. B., et McCarthy, W. F. (1989). "The Decade Census Program - A new approach for meeting the nation's needs for sub-national data." *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 351-355.
- Kish, L., Lovejoy, W., et Rackow, P. (1961), "A Multi-Stage Probability Sample for Traffic Surveys", *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 227-230.
- Kish, L. (1965), *Survey Sampling*. John Wiley & Sons, New York.
- Kish, L. (1979a), "Samples and Censuses", *International Statistical Review*, 47, pp. 99-109.
- Kish, L. (1979b), "Rolling Samples instead of Censuses", *Asian and Pacific Census Forum*, G(1), Août 1979, pp. 1-2, 12-13.
- Kish, L. (1981), "*Using Cumulated Rolling Samples to Integrate Census and Survey Operations of the Census Bureau*", Washington, D.C., U.S. Government Printing Office.
- Kish, L. (1983), "Data Collection for Details over Space and Time", T. Wright, (éd.), *Statistical Methods and the Improvement of Data Quality*, New York: Academic, pp. 72-84.
- Kish, L., et Verma, V (1983), Census plus Samples: Combined Uses and Designs, *Bulletin of the International Statistical Institute*, 50(1), pp. 66-82.

- Kish, L. (1984), "Timing of Surveys for Public Policy", *Australian Journal of Statistics*, pp. 1-12.
- Kish, L. et Verma, V., (1986), "Complete Censuses and Samples", *Journal of Official Statistics*, 2, pp. 381-93.
- Kish, L. (1987). *Statistical Research Design*, John Wiley & Sons, New York.
- Kish, L. (1990), "Recensement par étapes et échantillons avec renouvellement complet", *Techniques d'enquête*, 16 No 2, pp 67-86.
- Kish, L. (1997), "Periodic and Rolling Samples and Censuses", Bruce D. Spencer, éd., *Statistics and Public Policy*, chapitre 7, Clarendon Press, Oxford.
- Kish, L. (1998), "Space/Time Variations and Rolling Samples", *Journal of Official Statistics*, 14, 1, 1998, pp. 31-46.
- Kish, L. (1999). "Le cumul ou la combinaison d'enquêtes démographiques", *Techniques d'enquête*, 25, 2, pp. 147-158.
- Kish, L. (2001). "Combining Multi-population Surveys", *Journal of Statistical Planning and Inference*, to appear in 2001.
- Melnick, D. (1991). "The census of 2000 A. D. and beyond." *Reviews of Major Alternatives for the Census in the Year 2000*. U.S. Government Printing Office, Washington, D.C., August 1, 1991, pp. 60-74.
- National Academy of Sciences (1994). *Country People in the Information Age*. Duane L. Steffey et Norman M. Bradburn, éd. National Academy Press, Washington, D.C.
- National Academy of Sciences (1995). *Modernizing the U.S. Census*. Barry Edmonston et Charles Schultze, éd. National Academy Press, Washington, D.C.
- Purcell, N.J. et Kish, L. (1979). "Estimation for Small Domains," *Biometrics*, 35, pp. 365-384.
- Sawyer, T. C. (1993). "Rethinking the Census: Reconciling the Demands for Accuracy and Precision in the 21st Century." Présenté au Research Conference on Undercounted Ethnic Populations, 7 Mai, 1993.
- Tupek, A. R., Waite, P. J., et Cahoon, L. S. (1990). "Sample Expansion Plans for the Current Population Survey". *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 72-77.

Le présent article présente les résultats de travaux de recherche et d'analyse entrepris par les employés du Census Bureau. Il a fait l'objet, de la part du Bureau, d'un examen de portée plus limitée que celui auxquelles sont soumises les publications officielles. Le présent rapport est diffusé en vue de tenir les parties intéressées au courant des travaux de recherche en cours et d'en favoriser la discussion.