

Plans de sondage d'enquêtes téléphoniques auprès de ménages noirs aux États-Unis¹

KATHRYN M. INGLIS, ROBERTS M. GROVES, et STEVEN G. HEERINGA²

RÉSUMÉ

Le plan de sondage d'enquêtes téléphoniques à deux degrés avec règles de rejet élaboré par Waksberg en 1978 a été modifié pour améliorer l'efficacité des enquêtes téléphoniques auprès de la population noire aux États-Unis. L'application expérimentale d'autres plans de sondage a montré que: a) l'utilisation d'une stratification approximative fondée sur les noms des circonscriptions téléphoniques et des États; b) l'utilisation au premier degré de définitions de séries de 200 et 400 numéros consécutifs déterminant des grappes de grande taille; et c) l'application de règles de rejet fondées sur la race des ménages sont trois éléments dont la réunion améliore la précision relative d'un échantillon, étant donné des ressources fixes. Des modèles de coûts et des modèles d'erreur ont aussi été examinés pour simuler l'application d'autres plans de sondage.

MOTS CLÉS: Échantillons par CNH; enquêtes téléphoniques; échantillons de population rare.

1. INTRODUCTION

Les enquêtes effectuées auprès de populations rares et qui n'ont pas de base spéciale coûtent souvent plus cher par unité que celles qui visent une population complète. Quand la population rare est un petit sous-groupe d'une population facile à définir, l'échantillon de ce sous-groupe est souvent obtenu par filtrage de la population plus grande. Les enquêtes auprès de ménages de sous-groupes démographiques comme la population noire des États-Unis, procèdent en général de cette façon pour trouver les unités de sondage susceptibles d'être incluses dans l'échantillon; toutefois l'utilisation à grande échelle de techniques de filtrage pour constituer des échantillons de population rare coûte cher par interview. Ces dernières années, des méthodes de sondage par téléphone ont été proposées comme moyens peu coûteux pour échantillonner et interviewer des populations rares. Le coût d'une interview téléphonique est souvent inférieur à celui d'une interview en personne (Groves et Kahn 1979) et lorsqu'il faut absolument appliquer une méthode de filtrage pour déterminer les répondants susceptibles d'entrer dans un échantillon, l'efficacité des interviews téléphoniques en termes de coûts est encore plus marquée. Tout de même, les coûts de filtrage dans les enquêtes téléphoniques auprès de populations rares peuvent être élevés en chiffres absolus.

Dans le présent document, nous présentons diverses façons d'affiner la méthode de filtrage dans les enquêtes téléphoniques pour réduire les coûts tout en obtenant le degré de précision voulu. Pour cela, nous avons examiné divers plans de sondage qui pourraient s'appliquer à des enquêtes téléphoniques auprès de la population noire des États-Unis.

Les essais d'enquête téléphonique décrits ici ont été effectués dans le cadre d'un sondage d'opinion mené auprès des Noirs au moment de la campagne électorale de 1984 pour la présidence des États-Unis afin de connaître leurs attitudes politiques et leurs intentions de vote.

¹ Une version révisée d'un article qui a été présenté au congrès de 1985 de l'American Statistical Association. La recherche a été partiellement subventionnée par l'U.S. Bureau of the Census et le Survey Research Center. La discussion ne représente par nécessairement les opinions de ces organismes-là.

² Kathryn M. Inglis, McNair Anderson and Associates, Australie. Robert M. Groves et Steven G. Heeringa, Survey Research Center, University of Michigan, Ann Arbor, Michigan, 48106-1248, États-Unis.

L'utilisation de techniques d'échantillonnage et d'interview par téléphone exclut de l'enquête les Noirs faisant partie de ménages qui n'ont pas le téléphone (soit 15% environ de toute la population noire aux États-Unis). Dans la plupart des cas, les personnes non abonnées au téléphone sont plus pauvres et plus jeunes que celles qui vivent dans les ménages abonnés au téléphone (Thornberry et Massey 1983). Dans la mesure où il est possible que les Noirs non abonnés au téléphone n'aient ni les mêmes attitudes politiques ni les mêmes intentions de vote que ceux qui sont abonnés, les estimations produites à partir des résultats de l'enquête risquent de différer des vrais paramètres de l'ensemble de la population noire des États-Unis. Sans chercher à minimiser l'erreur de non couverture associée aux enquêtes téléphoniques auprès de la population noire, nous nous sommes principalement attachés aux différences au niveau de l'efficacité sur le plan des coûts et de l'erreur d'échantillonnage pouvant résulter de l'utilisation d'approches alternatives dans la formation des échantillons de ménages noirs pour des enquêtes téléphoniques.

Les plans de sondage pour enquêtes téléphoniques présentés ici ont été élaborés à partir d'un plan défini par Waksberg (1978). Ce plan de sondage par composition de numéros de téléphone au hasard (CNH), communément appelé méthode de Waksberg-Mitofsky, est un plan d'échantillonnage de numéros de téléphone par grappes à deux degrés. Les numéros de téléphone aux États-Unis comprennent 10 chiffres, soit un code régional à 3 chiffres, un code ou indicatif de central à trois chiffres et un numéro de ligne à 4 chiffres de 0000 à 9999, par exemple 313-764-4424. À la première étape, un échantillon stratifié de numéros de téléphone à dix chiffres est produit au hasard et chaque "numéro primaire" ainsi produit est lié à un bloc de 100 numéros consécutifs (par exemple le numéro 313-764-4424 serait lié à la série de 100 numéros consécutifs 313-764-4400 à 313-764-4499). Si, dans le cas des enquêtes-ménages, il se trouve que le numéro primaire est un numéro de téléphone de ménage valide, alors la série de 100 numéros de téléphone consécutifs dont il fait partie est retenue à la première étape pour une deuxième étape d'échantillonnage. Sinon, sa série de 100 numéros consécutifs est rejetée. Par conséquent, la probabilité de sélection d'une série de 100 numéros consécutifs à la première étape est proportionnelle au nombre de numéros de téléphone de ménage valides dans cette série. À la deuxième étape de l'échantillonnage, des numéros de téléphone de ménage valides sont choisis en nombre égal dans chacune des séries de 100 numéros consécutifs retenues à la première étape. Par conséquent, l'échantillonnage des ménages, à la deuxième étape, est un échantillonnage avec probabilité conditionnelle de sélection inversement proportionnelle au nombre de numéros de téléphone de ménage valides dans chaque série de 100 numéros consécutifs. Le plan produit ainsi un échantillon de numéros de téléphone de ménage avec équiprobabilité de sélection ("epsem sampling") et regroupe ces numéros par grappes de sorte que la proportion du nombre total de numéros de téléphone sélectionnés qui rejoignent un ménage est plus grande que celle qui aurait été obtenue au moyen d'un échantillonnage stratifié par CNH. Par souci de clarté dans l'exposé qui va suivre, nous appellerons les banques de 100 numéros de téléphone consécutifs, unités de sondage du premier degré ou unités primaires d'échantillonnage (UPE) du plan de sondage par CNH à deux degrés. Le terme "grappe" est réservé pour désigner l'ensemble comprenant un nombre fixé de numéros de téléphone de ménage valides choisis à partir des UPE à la deuxième étape du plan de sondage.

Dans la présente étude, les modifications apportées au plan de sondage pour réduire les coûts de filtrage prennent trois formes: a) stratification des circonscriptions téléphoniques selon la proportion de Noirs et prélèvement non proportionnel (c'est-à-dire d'une proportion plus élevée) d'une partie de l'échantillon dans les strates à forte densité de population noire; b) utilisation de règles de rejet à deux degrés fondées sur le fait que le numéro de téléphone composé est un numéro résidentiel ou non et sur la race du ménage; et c) augmentation de la taille des UPE (de 100 numéros consécutifs à 200 et 400).

La stratification de la population des ménages abonnés au téléphone selon la race a pour objet d'isoler les circonscriptions (zones de rattachement à un central) à forte proportion d'abonnés de race noire. On applique ensuite à ces strates des fractions d'échantillonnage plus grandes que celles qui sont appliquées aux strates qui ont une plus faible proportion de ménages noirs. Selon ce plan de sondage à répartition non proportionnelle, le nombre total de ménages qu'il faut rejoindre pour obtenir une interview avec un ménage noir admissible est plus petit que le nombre de ménages qu'il faudrait rejoindre si l'on utilisait un échantillon "epsem" de la population des ménages. Les coûts de filtrage qu'entraîne la constitution d'un échantillon de ménages noirs sont ainsi diminués. Dans les échantillons d'enquêtes téléphoniques, l'unité géographique de base est la circonscription correspondant à un central téléphonique et à laquelle un ou plus d'un indicatif à trois chiffres est attribué. En général, le compte des abonnés n'est pas ventilé selon les caractéristiques raciales. Il faut par conséquent utiliser des indicateurs approximatifs des circonscriptions à forte densité de population noire. Les expériences décrites dans le présent document avaient pour objet notamment d'examiner la valeur de ces indicateurs d'approximation.

Blair et Czaja (1982) ont élaboré une variante du plan de sondage par CNH de Waksberg-Mitofsky qui intègre des règles de rejet à deux degrés fondées aussi bien sur le fait que le numéro composé est un numéro résidentiel ou non et sur la race du ménage. Appliquée à la population noire, cette méthode inclut à la première étape seulement des séries de 100 numéros consécutifs dont le numéro primaire a été attribué à un ménage noir et prélève ensuite dans chacune des UPE un nombre fixé d'avance de numéros de téléphone de ménages noirs. Dans une enquête nationale par sondage aux États-Unis, Blair et Czaja ont constaté qu'en utilisant ce plan, le pourcentage de ménages noirs rejoints à l'aide de tous les numéros de téléphone choisis augmentait à 25 pour cent à la deuxième étape, tandis qu'il était de 9 pour cent à la première étape. Étant donné les probabilités de sélection qui se compensent dans les deux étapes, ce plan de sondage "epsem" réduit considérablement le travail de filtrage requis pour obtenir un échantillon de taille donnée de ménages noirs. Une variante analogue des règles de rejet utilisées dans le plan de sondage à deux degrés de Waksberg-Mitofsky a été utilisée dans les expériences décrites dans le présent document.

Dans le plan de sondage de Blair et Czaja, certaines séries de 100 numéros consécutifs de la première étape ne contenaient pas assez de numéros de téléphone de ménages noirs pour produire le nombre d'éléments par grappe (10 dans ce cas-là) qu'il faut pour constituer un échantillon "epsem" de ménages noirs. En outre, ce plan comporte des coûts assez élevés de filtrage au premier degré de sélection; il faut composer plus de 44 numéros primaires pour rejoindre un ménage noir. La solution commune à ces deux problèmes est d'augmenter la taille de l'UPE et de choisir plus d'éléments dans les UPE à la deuxième étape. Les analyses présentées ici ont utilisé des unités de sondage du premier degré de 100, 200 et 400 numéros consécutifs chacune. L'extension de la définition des UPE pour qu'elles comprennent plus de numéros que la norme habituelle de 100 numéros consécutifs est suggérée par les observations sur la répartition des numéros de téléphone à l'intérieur d'un indicatif donné. Voici les principales constantes qui semblent se dégager de nos observations: 1) presque tous les numéros de ménages dont le numéro de téléphone commence par un indicatif donné desservent des unités situées à l'intérieur des limites géographiques de la circonscription téléphonique correspondant à l'indicatif en question; 2) il y a peu de regroupements géographiques par grappes des numéros attribués à l'intérieur des limites géographiques des circonscriptions téléphoniques (c'est-à-dire qu'en général deux ménages voisins n'ont pas deux numéros consécutifs et n'ont pas nécessairement deux numéros comportant le même indicatif); et 3) le pourcentage des numéros rejoignant des ménages varie plus d'une série de 1,000 numéros consécutifs à une autre qu'entre les séries de 100 numéros consécutifs d'une même série de 1,000 numéros consécutifs. Ces constatations sont le fruit de plusieurs années d'enquêtes

téléphoniques effectuées par le Survey Research Center. Elles donnent à penser que l'extension de la définition des UPE de 100 numéros consécutifs à un nombre plus élevé pourrait nous permettre d'utiliser des grappes de numéros secondaires plus grandes sans beaucoup réduire la proportion de ces numéros qui rejoignent des ménages noirs.

2. L'ÉTUDE PILOTE

Dans deux expériences intégrées effectuées dans le cadre d'une enquête pilote, plusieurs plans de sondage différents ont été testés. Un des objectifs de l'étude pilote était de déterminer dans quelle mesure on peut produire des séries de numéros de téléphone à forte densité de numéros d'abonnés noirs au moyen d'une stratification fondée sur les régions administratives et d'une correspondance approximative entre les limites des régions administratives et celles des circonscriptions téléphoniques. Pour cela, trois strates de circonscriptions ont été définies:

- 1) "À forte densité" – Circonscriptions téléphoniques correspondant aux villes centrales des grandes régions métropolitaines statistiques normalisées (par exemple la ville de Chicago pour la RMSN de Chicago). Cette définition était fondée sur le nom des circonscriptions de ces régions.
- 2) "À densité moyenne" – Toutes les autres circonscriptions dans certains États choisis du sud des États-Unis (Virginie, Caroline du Nord, Caroline du Sud, Floride, Georgie, Alabama, Mississippi et Louisiane). La très grande majorité des circonscriptions de ces États sont comprises dans un seul État; celles qui desservaient deux États en même temps ont été associées à l'État dont le nom apparaît dans le nom de la circonscription.
- 3) "À faible densité" – Toutes les autres circonscriptions des États limitrophes des États-Unis.

Un échantillon avec équiprobabilité de sélection de 1,400 combinaisons à six chiffres dont trois pour l'indicatif régional et trois pour l'indicatif du central a ensuite été prélevé systématiquement à partir des 34,389 combinaisons de ce genre qui sont inscrites comme étant en service dans une base qu'on peut acheter de la société American Telephone & Telegraph (AT&T). Des numéros aléatoires à 4 chiffres ont été ajoutés à chacun des radicaux à six chiffres choisis pour produire un échantillon de 1,400 numéros primaires à dix chiffres.

Les résultats de l'étude pilote ont montré que la proportion de numéros de téléphone rejoignant des ménages noirs variait beaucoup selon la strate. On a constaté qu'il fallait faire six fois plus de filtrage pour trouver un ménage noir dans la strate à faible densité que dans la strate à forte densité. (Ce résultat a été confirmé avec plus de précision dans l'étude de production, dont nous parlerons dans la prochaine section).

Un autre objectif de l'étude pilote était de tester l'utilisation de règles de rejet fondées sur la composition raciale des ménages correspondant aux numéros de téléphone de l'échantillon tiré des UPE de différentes tailles et sur la validité/non validité de ces numéros (suivant qu'ils rejoignent ou non un ménage). Pour augmenter la précision des analyses relatives à cet objectif, 500 autres numéros primaires ont été choisis dans les strates à forte et à moyenne densité. Les 1,900 numéros primaires de l'échantillon combiné de l'étude pilote ainsi prélevés ont ensuite été composés pour repérer les ménages noirs. Quand un numéro primaire échantillonné rejoignait un ménage noir, il correspondait en même temps à trois UPE différentes. Comme on peut le voir au tableau 1, chaque numéro individuel peut être considéré comme faisant partie en même temps d'une série de 100 numéros consécutifs, d'une série de 200 numéros consécutifs et d'une série de 400 numéros consécutifs. Par exemple, le numéro 313-764-4424 fait partie de la série des 100 numéros consécutifs 4400-4499, de la série des 200 numéros consécutifs 4400-4599 et de la série des 400 numéros consécutifs 4400-4799.

Pour tester la possibilité d'augmenter la taille des UPE, les responsables de l'étude pilote ont prélevé des numéros secondaires dans chacune de ces trois séries. La taille des grappes de ménages noirs du second degré a été fixée à 3 pour les grappes constituées à partir de chaque série de 100 numéros consécutifs associée à un numéro primaire, à 6 pour les grappes constituées à partir de chaque série de 200 numéros consécutifs et à 9 pour les grappes constituées à partir de chaque série de 400 numéros consécutifs. Au premier comme au second degré de sélection, on supposait, si la race du ménage n'était pas connue, qu'il ne s'agissait pas d'un ménage noir.

Le tableau 1 présente la répartition des numéros secondaires selon le type d'UPE et la strate. Ce qui nous intéresse le plus, c'est la proportion de numéros secondaires attribués aux ménages noirs pour les différentes définitions de l'UPE. Pour les séries de 100 numéros consécutifs, .134 de tous les numéros secondaires sont des numéros rejoignant des ménages noirs. Cela entraîne que .223 de tous les ménages échantillonnés étaient noirs, comparative-ment à la proportion de .25 obtenue par Blair et Czaja. Pour les UPE de 200 numéros consécutifs, .124 de tous les numéros secondaires étaient des numéros rejoignant des ménages noirs, tandis que pour les UPE de 400 numéros consécutifs, .115 de tous les numéros de téléphone prélevés à la deuxième étape étaient attribués à des ménages noirs. L'écart entre ces proportions est toujours inférieur à l'erreur d'échantillonnage (l'écart-type de chaque estimation est d'au moins .02). Cela signifie qu'en portant de 100 à 400 numéros consécutifs la longueur des UPE, on ne diminue pas sensiblement la proportion des numéros rejoignant des ménages noirs (numéros admissibles). Ces taux entraînent qu'étant donné que les UPE de 100 numéros consécutifs permettent d'obtenir en moyenne au second degré des grappes de 13 ou 14 ménages noirs, les séries de 400 numéros consécutifs devraient permettre en moyenne d'obtenir des grappes de 46 ménages noirs. La possibilité d'augmenter la taille des grappes de ménages noirs au second degré d'échantillonnage permet aux chercheurs de réduire considérablement les coûts de filtrage.

Le tableau 1 permet également de comparer les proportions de numéros secondaires admissibles échantillonnés à partir des trois différentes strates utilisées dans l'étude pilote et selon les trois définitions des UPE retenues. Pour toutes les définitions des UPE retenues (100, 200 et 400 numéros consécutifs), on obtient le même résultat - les circonscriptions téléphoniques des grandes RMSN auxquelles correspond la strate à forte densité de population noire produisent des taux d'admissibilité presque deux fois plus élevés que le taux obtenu pour l'ensemble de la population (.21 contre .12 ou .13). La strate à densité moyenne, comprenant les ménages des circonscriptions téléphoniques hors des RMSN de certains États choisis du sud des États-Unis, produit des taux d'admissibilité inférieurs au taux national (soit des taux variant entre .08 et .10). La strate à faible densité, correspondant aux autres États limitrophes des États-Unis, produit également des taux inférieurs à la moyenne (oscillant entre .07 et .085). Comme la strate à forte densité couvre environ 36 pour cent des ménages noirs abonnés au téléphone, la stratification choisie conjuguée avec la répartition non proportionnelle de l'échantillon au premier degré est un moyen efficace pour réduire les coûts de filtrage.

3. L'ÉTUDE DE PRODUCTION

L'étude de production a utilisé le plan de stratification élaboré et testé dans l'étude pilote. Un échantillon de 11,223 numéros primaires répartis de façon non proportionnelle a été tiré des trois strates selon la densité de la population noire à l'aide de fractions d'échantillonnage dans des proportions de 3, 2 et 1 (3 pour la strate à forte densité, 2 pour la strate à densité moyenne et 1 pour la strate à faible densité). Même si l'étude pilote n'a pas révélé

Tableau 1
Étude pilote
Répartition des numéros secondaires choisis dans les séries de
100, 200 et 400 numéros consécutifs par strate

Strate et répartition	Proportion de tous les numéros choisis		
	Séries de 100	Séries de 200	Séries de 400*
Strate à forte densité de Noirs			
Ménages noirs	.205	.201	.214
Ne connaît pas la race	.028	.029	.032
Ménages non noirs	.316	.279	.275
Numéros non résidentiels/non valides	.451	.491	.479
Nombre de cas	(395)	(806)	(1163)
Strate à densité moyenne de Noirs			
Ménages noirs	.104	.080	.076
Ne connaît pas la race	.030	.018	.020
Ménages non noirs	.494	.443	.420
Numéros non résidentiels/non valides	.372	.459	.484
Nombre de cas	(231)	(560)	(878)
Strate à faible densité de Noirs			
Ménages noirs	.085	.084	.069
Ne connaît pas la race	.014	.028	.027
Ménages non noirs	.532	.577	.607
Numéros non résidentiels/non valides	.369	.311	.297
Nombre de cas	(141)	(286)	(491)
Total			
Ménages noirs	.134	.124	.115
Ne connaît pas la race	.024	.025	.026
Ménages non noirs	.442	.431	.448
Numéros non résidentiels/non valides	.400	.420	.411
Nombre de cas	(767)	(1652)	(2532)

* Estimation pondérée pour tenir compte du fait que les 9 numéros secondaires des grappes obtenues à partir des séries de 400 numéros consécutifs ne sont pas répartis proportionnellement entre chacune des 4 tranches de 100 numéros consécutifs.

de différence appréciable entre les taux de validité pour les UPE de 200 et de 400 numéros consécutifs, on a opté pour la prudence en choisissant les séries de 200 numéros consécutifs dans l'étude de production. La taille attendue des grappes au second degré a été fixée à 5.5 ménages noirs (sans compter le numéro primaire). Les règles de rejet au premier et au second degré appliquées dans la version modifiée du plan de sondage à deux degrés de Waksberg-Mitofsky ayant servi dans l'étude pilote ont aussi été appliquées dans l'étude de production. Comme des tailles d'échantillon beaucoup plus grandes ont été utilisées dans l'étude de production, on peut répondre aux questions concernant la précision et l'efficacité relative du plan utilisé avec plus d'assurance.

Le tableau 2 présente les résultats du filtrage des numéros primaires et secondaires effectué dans l'étude de production. D'après les taux estimatifs pondérés non biaisés obtenus suivant un plan de sondage "epsem" par CNH à deux degrés, 13 pour cent de tous les numéros

secondaires étaient des numéros de ménages noirs (l'écart-type pour cette estimation est de .6 pour cent). Cela est très proche du taux d'admissibilité des numéros secondaires de 12 pour cent enregistré dans l'étude pilote. La comparaison des résultats obtenus au premier degré de sélection avec les résultats obtenus au second degré de sélection montre les gains importants qu'on peut réaliser en utilisant un plan de sondage à deux degrés pour faire des enquêtes téléphoniques auprès des ménages noirs. Les gains à tirer d'un plan de sondage à deux degrés sont le plus marqués dans la strate à faible densité de population noire, où la proportion de ménages rejoints augmente de presque neuf fois entre le premier et le second degré (de .011 à .090). Dans la strate à forte densité, l'augmentation est plus proche du double (.072 à .190). Quant au cas où le plan de sondage utilisé est un plan de sondage à répartition non proportionnelle, les proportions non pondérées de ménages noirs rejoints aux deux degrés de sélection sont de 3 pour cent (au premier degré) et de 15 pour cent (au second degré). Si l'on compare ces chiffres aux estimations obtenues selon le plan de sondage "epsem" (c'est-à-dire à 2 et à 13 pour cent), on constate que les coûts de filtrage sont moins élevés dans le premier cas.

Tableau 2
Étude de production
Répartition des numéros choisis selon la strate

Strate et répartition	Primaires	Secondaires
Strate à forte densité de Noirs		
Ménages noirs	.072	.190
Ne connaît pas la race	.035	.027
Ménages non noirs	.219	.352
Numéros non résidentiels/non valides	.674	.431
Nombre de cas	(3,128)	(6,671)
Strate à densité moyenne de Noirs		
Ménages noirs	.032	.141
Ne connaît pas la race	.020	.018
Ménages non noirs	.188	.469
Numéros non résidentiels/non valides	.760	.372
Nombre de cas	(1,879)	(2,375)
Strate à faible densité de Noirs		
Ménages noirs	.011	.090
Ne connaît pas la race	.019	.023
Ménages non noirs	.199	.505
Numéros non résidentiels/non valides	.771	.382
Nombre de cas	(6,116)	(3,987)
Plan de sondage "epsem"*		
Ménages noirs	.021	.129
Ne connaît pas la race	.021	.023
Ménages non noirs	.200	.454
Numéros non résidentiels/non valides	.758	.394
Proportion de ménages noirs selon un plan de sondage à répartition non proportionnelle		
	.031	.150
Nombre de cas	(11,123)	(13,033)

* Estimations pondérées des taux obtenus selon un "plan de sondage epsem". Les coefficients de pondération compensent le fait que des fractions de sondage non proportionnelles ont été utilisées pour choisir l'échantillon de l'étude de production à partir des trois strates de densités différentes.

Comme dans l'étude pilote, le pourcentage de ménages noirs rejoints est différent pour chacune des trois strates, encore que la distinction entre la strate à densité moyenne et la strate à faible densité soit la plus intéressante. Le pourcentage de ménages noirs rejoints à l'aide des numéros secondaires varie approximativement dans des proportions de 2, 1.5 et 1 entre les trois strates. Les trois strates diffèrent également entre elles par la proportion, dans chacune, de numéros de téléphone attribués à des résidences. La strate à forte densité de population noire a proportionnellement plus de numéros secondaires attribués à des unités non résidentielles, ce qui reflète probablement le fait qu'une plus grande partie des circonscriptions téléphoniques de cette strate est urbaine.

Chaque UPE de 200 numéros consécutifs peut être considérée comme deux moitiés d'UPE de 100 numéros consécutifs chacune. Le tableau 3 montre que la proportion de numéros non résidentiels qu'on trouve dans les 100 premiers numéros consécutifs des UPE dans lesquelles le numéro primaire tombait est plus faible que celle qu'on trouve dans l'autre moitié (.378 comparativement à .409), mais cette différence n'est pas statistiquement significative au niveau de .05 (écart-type de .02 environ). De même, la proportion de ménages noirs rejoints à l'aide des numéros secondaires est un peu plus élevée lorsque les numéros secondaires tombaient dans la première moitié de 100 numéros consécutifs du numéro primaire que lorsqu'ils tombaient dans la série de 100 numéros consécutifs adjacente (.133 comparativement à .125). Mais il est peu probable qu'on observe la même différence dans la plupart des cas où cette expérience est répétée. Le tableau 3 donne un autre point de vue sur les résultats du tableau 2; il montre que la réduction observée entre la proportion des ménages admissibles obtenue à partir des séries de 100 numéros consécutifs dans lesquelles le numéro primaire tombe et la proportion obtenue à partir des séries de 100 numéros consécutifs adjacentes est très négligeable.

Le taux d'admissibilité moyen obtenu sur l'ensemble des UPE, qui correspond à la proportion de ménages noirs rejoints, ne devrait pas être le seul critère d'évaluation des plans de sondage. Pour appliquer un plan de sondage "epsem" stratifié, chaque UPE retenue doit avoir un nombre suffisant de ménages noirs pour permettre de fixer le nombre de ménages noirs souhaité au second degré. Par conséquent, la distribution de la proportion des numéros admissibles sur l'ensemble des UPE est également importante. Les histogrammes des figures 1, 2 et 3 montrent la distribution de la proportion des ménages noirs par strate sur l'ensemble des UPE. La stabilité des trois distributions varie parce qu'il y a environ quatre fois plus d'UPE dans la strate à forte densité que dans les deux autres (224 UPE dans la strate à forte densité contre 60 environ dans la strate à densité moyenne et la strate à faible densité). Par ailleurs, la forme de la courbe de distribution semble très différente selon la strate.

Tableau 3

Étude de production
Répartition des numéros secondaires selon qu'ils tombent dans la même série
de 100 numéros consécutifs que le numéro primaire ou dans la série de
100 numéros consécutifs adjacente

Situation	Répartition	
	Même série de 100 numéros consécutifs que le numéro primaire	Série de 100 numéros consécutifs adjacente
Ménages noirs	.133	.125
Ne connaît pas la race	.024	.022
Ménages non noirs	.465	.444
Numéros non résidentiels/non valides	.378	.409
Nombre de cas	(6,522)	(6,511)

Pourcentage de grappes

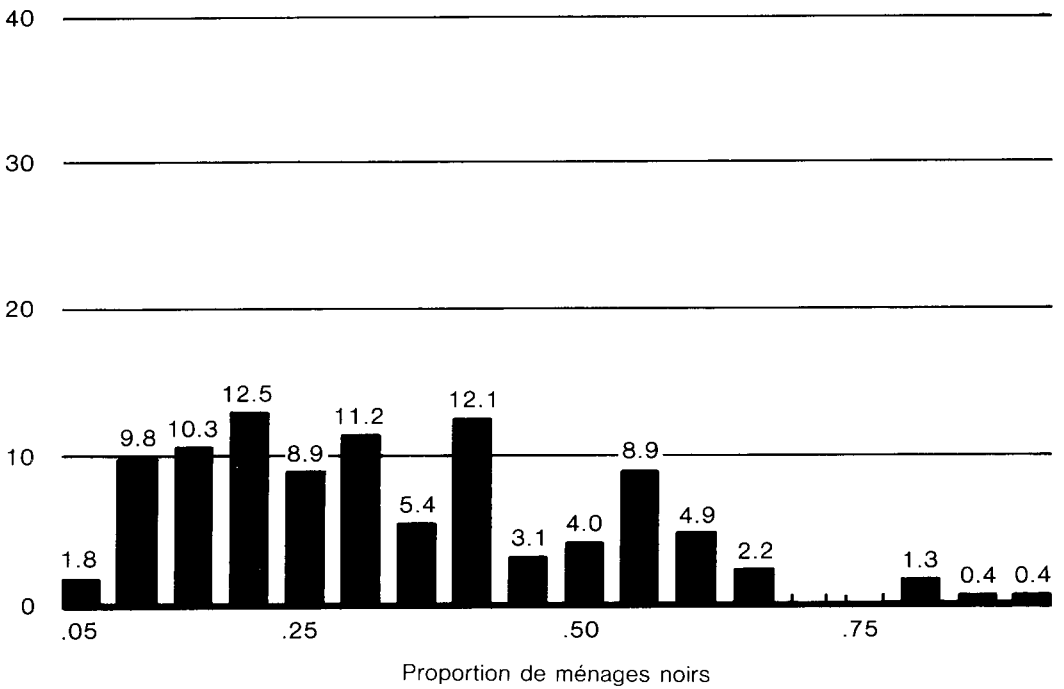


Figure 1. Pourcentage de grappes à forte densité selon la proportion de ménages noirs

Pourcentage de grappes

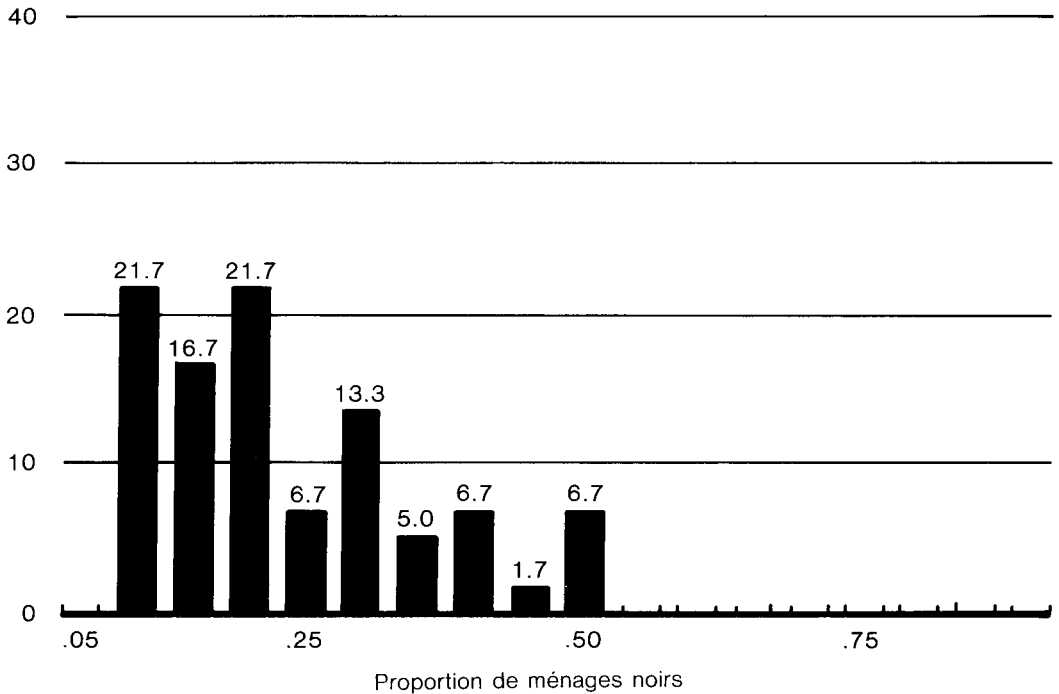


Figure 2. Pourcentage de grappes à densité moyenne selon la proportion de ménages noirs

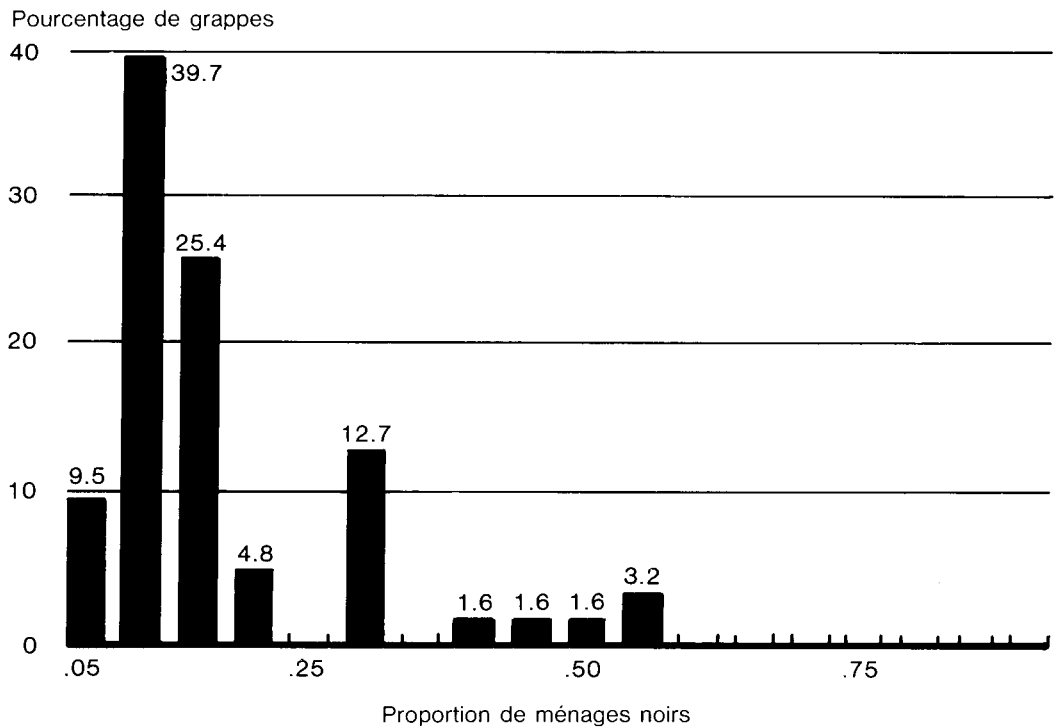


Figure 3. Pourcentage de grappes à faible densité selon la proportion de ménages noirs

La distribution des grappes dans les strates à faible densité et à densité moyenne est très asymétrique, 60 pour cent des UPE de la strate à densité moyenne et 75 pour cent des UPE de la strate à densité moyenne ayant de 5 à 20 pour cent de ménages noirs. Ces taux d'admissibilité correspondent à un maximum de 10 à 40 ménages noirs pour les UPE de 200 numéros consécutifs prélevées dans la strate à faible densité et la strate à densité moyenne. Dans l'étude de production, la strate à faible densité contenait plusieurs UPE qui n'auraient pas permis de former des grappes de cette taille. (On a estimé que 6 des 63 UPE de cette strate avaient moins de 10 ménages noirs.) La distribution des grappes dans la strate à forte densité est beaucoup plus uniforme. (On a estimé que seulement 4 des 224 UPE de cette strate avaient moins de 10 ménages noirs.)

Ces distributions du pourcentage de ménages noirs par UPE méritent plus d'explications. Compte tenu de ce que nous savons déjà de l'attribution des numéros résidentiels aux banques de numéros disponibles, il n'y a pas lieu de croire qu'à l'intérieur d'une circonscription téléphonique donnée (ou d'un indicatif de central donné), il y ait une tendance générale à attribuer des zones résidentielles différentes à des séries de 100 numéros consécutifs différentes. Cela signifie qu'à l'intérieur d'une circonscription desservant aussi bien des ménages noirs que des ménages non noirs, tout semble indiquer qu'il faut presque nécessairement faire l'hypothèse que les numéros sont attribués sans égard pour la race de l'abonné. Autrement dit, à moins que les circonscriptions ne soient subdivisées en zones rattachées à des noeuds de câbles correspondant à des zones résidentielles de ménages noirs, il n'y a pas de raison *a priori* de supposer que les grappes de ménages noirs dans les séries de 200 numéros consécutifs sont grandes. Suivant ce raisonnement, la distribution plus uniforme des grappes dans la strate à forte densité reflète, croyons-nous, le fait que la proportion de ménages noirs a tendance à varier entre les populations d'abonnés au téléphone des différentes circonscriptions téléphoniques composant la strate.

4. CARACTÉRISTIQUES DE LA VARIANCE D'ÉCHANTILLONNAGE

Pour réduire les coûts de l'échantillonnage de ménages noirs par CNH, il est avantageux d'utiliser de grandes grappes de ménages par UPE (c'est-à-dire d'utiliser moins d'UPE pour une taille d'échantillon donnée) et de tirer proportionnellement plus d'UPE des strates composées des circonscriptions où la proportion de ménages noirs abonnés au téléphone varie. Bien que l'utilisation de plus grosses grappes et la répartition non proportionnelle de l'échantillon améliorent l'efficacité en termes de coûts, la précision globale de l'échantillon est diminuée par l'effet de l'utilisation de plus grosses grappes et par les effets du plan de sondage rendus plus marqués par la nécessité d'effectuer une pondération non optimale pour tenir compte des probabilités inégales de sélection des ménages à partir des trois strates selon la densité de la population noire. Kish (1976) a décrit, les effets augmentés du plan de sondage des estimations dans les cas où la pondération n'est pas optimale. Nous montrons dans les paragraphes qui suivent comment l'utilisation de grappes influe les effets du plan de sondage dans la version modifiée de la méthode d'échantillonnage par CNH utilisée dans la présente étude.

Toutes choses étant égales par ailleurs, plus on choisit d'éléments par UPE, plus l'effet du plan de sondage est grand (l'effet du plan de sondage est le ratio de la variance d'échantillonnage d'un plan de sondage donné sur celle d'un échantillon aléatoire simple comprenant le même nombre d'éléments). Pour calculer l'effet du plan de sondage, le modèle le plus souvent utilisé est $Deff = 1 + \rho (b - 1)$, où $Deff$ ("Design effect") est l'effet du plan de sondage, ρ , la corrélation intragrappe pour la statistique considérée et b , le nombre d'éléments de l'échantillon par UPE. D'autres études ont montré que, pour beaucoup de variables portant sur l'ensemble de la population aux États-Unis, la corrélation intragrappe observée avec les séries de 100 numéros consécutifs est la plupart du temps inférieure à celle qu'on trouve en général entre les grappes d'échantillons probabilistes aréolaires (voir Groves 1978). Il est possible que cela ne soit pas le cas pour la population noire avec les séries de 100 numéros consécutifs; par ailleurs aucune estimation empirique de la corrélation intragrappe n'a encore été faite avec des séries de 200 numéros consécutifs. L'hypothèse formulée avant d'estimer les erreurs d'échantillonnage était que la corrélation intragrappe devrait avoir la même valeur pour les séries de 100 comme de 200 numéros consécutifs. Cette hypothèse reflète la conception que nous avons formulée plus haut de l'attribution des numéros de téléphone à l'intérieur des circonscriptions téléphoniques.

Suivant les erreurs d'échantillonnage dont la valeur a été estimée à partir d'un ensemble de données de l'étude de production, la valeur moyenne de l'effet du plan de sondage obtenue pour sept statistiques choisies a été de 1.28 avec les séries de 100 numéros consécutifs et de 1.30 avec les séries de 200 numéros consécutifs. La valeur moyenne de l'effet du plan de sondage obtenue pour les séries de 100 numéros consécutifs a été estimée à partir des seuls cas qui tombaient dans la même série de 100 numéros consécutifs que le numéro primaire, tandis que tous les cas ont été utilisés pour calculer la valeur moyenne de l'effet du plan de sondage pour les séries de 200 numéros consécutifs. Ainsi, la taille moyenne des grappes des interviews complétées est de 2.0 dans le cas des séries de 100 numéros consécutifs (coefficient de variation de .043) et de 3.4 pour les séries de 200 numéros consécutifs (coefficient de variation de .029). Ces effets du plan de sondage reflètent la stratification, le regroupement par grappes et la pondération prévues dans le plan de sondage et aussi le fait que la taille des grappes prélevées dans les séries de 100 numéros consécutifs varie plus. (Une des règles de rejet obligeait à prélever le même nombre de ménages noirs dans les séries de 200 numéros consécutifs, mais pas nécessairement le même nombre dans les séries de 100 numéros consécutifs.) Étant donné que les effets du plan de sondage obtenues avec les séries de 100 et les séries de 200 sont très proches l'une de l'autre (1.28 et 1.30), il semble que l'élément

Tableau 4
Étude de production
Corrélations intragrappe synthétiques selon la strate et selon que les grappes sont tirées
des séries de 100 ou de 200 numéros consécutifs pour sept statistiques choisies

Statistique	Corrélation intragrappe synthétique*					
	Strate à forte densité de population noire		Strate à densité moyenne de population noire		Strate à faible densité de population noire	
	Séries de 100	Séries de 200	Séries de 100	Séries de 200	Séries de 100	Séries de 200
Proportion très satisfaite de la vie en général	.021	-.002	-.172	-.042	-.238	-.116
Proportion qui estime être dans une meilleure situation financière que l'année précédente	.113	.075	.094	.069	.206	.049
Proportion qui votera pour Mondale	.189	.021	.086	-.087	-.436	-.046
Proportion qui va à l'église	.013	.017	-.009	-.078	.035	-.110
Proportion ayant passé toute sa vie dans la même ville	-.078	.001	.058	.114	.221	.248
Proportion qui a voté aux élections présidentielles de 1980	-.045	-.035	-.101	-.013	.364	.356
Proportion qui pense que Reagan sera élu président	-.045	-.045	-.545	-.078	.124	-.105
Moyenne	.024	.005	-.084	-.016	.039	.039

* Ces estimations ne sont pas pondérées.

déterminant de la variance d'échantillonnage soit la pondération non optimale qu'il faut effectuer pour tenir compte du fait que l'échantillon n'est pas réparti de façon proportionnelle, tandis que l'augmentation de la taille des UPE de 100 à 200 numéros consécutifs ne fait perdre que très peu de précision.

Le tableau 4 présente les résultats du calcul des corrélations synthétiques intragrappe par strate pour les sept statistiques de l'enquête utilisées pour estimer l'effet moyen du plan de sondage. Les estimations des corrélations synthétiques intragrappe ont été obtenues à partir de la valeur de l'effet du plan du sondage à l'aide du modèle de Kish où $Rho = (Deff - 1) / (b - 1)$ et n'ont pas été pondérées pour qu'il n'y ait pas d'effet de confusion de la pondération sur les estimations synthétiques. Les estimations du tableau ont tendance à être instables à cause du petit nombre de grappes prélevées dans chaque strate, de l'assez faible taille moyenne des grappes une fois les interviews complétées et du coefficient de variation associé. Ces caractéristiques du plan de sondage rendent plus difficile pour nous de tirer des conclusions à propos de l'effet du regroupement par grappes dans les séries de 100 et de 200 numéros consécutifs. En général, les estimations de la corrélation intragrappe sont un peu plus élevées dans les séries de 100 que dans les séries de 200. Nous croyons que cela reflète plus l'instabilité de la corrélation synthétique estimative qu'une différence réelle au niveau de l'effet du regroupement par grappes et que ces estimations ne donnent pas beaucoup d'évidence que la corrélation intragrappe soit différente selon qu'on la calcule pour des séries de 100 ou de 200 numéros consécutifs.

5. CARACTÉRISTIQUES OPTIMALES DU PLAN DE SONDAGE

Les sections précédentes portaient sur l'effet de l'utilisation de différents plans de sondage sur l'efficacité en termes de coûts et sur la variance d'échantillonnage. Dans la plupart des enquêtes par sondage, c'est à l'étape de l'élaboration du plan de sondage qu'on compare les coûts d'enquête et les erreurs pour chercher à déterminer quelles devraient être les caractéristiques "optimales" de l'enquête. Suivant cette façon de procéder, on tente d'élaborer le plan de sondage qui permettra d'obtenir la plus petite variance possible pour un ensemble donné de ressources à consacrer à l'enquête. Vu la nature de la présente étude, la recherche du choix optimal portera avant tout sur les deux caractéristiques suivantes du plan de sondage: a) nombre d'éléments par UPE et b) répartition de l'échantillon entre les trois strates définies selon la densité de la population noire.

Pour déterminer la taille optimale des grappes, nous avons utilisé le modèle de coûts totaux suivant: $C = C_o + C_a a + C_b ab$, où C_o représente les coûts fixes, où C_a représente le coût d'échantillonnage et de filtrage par grappe et a le nombre de grappes choisies et où C_b représente le coût d'échantillonnage, de filtrage et d'interview associé à chaque interview obtenue et b le nombre d'interviews obtenues dans chaque grappe. Comme la proportion de ménages noirs varie d'une strate à l'autre, la valeur des paramètres C_a et C_b varie aussi selon la strate (voir tableau 5). La taille optimale des grappes est donnée par $\sqrt{C_a(1 - \rho)/(C_b\rho)}$ (Kish 1965). Le tableau 5 présente aussi les résultats des estimations que nous avons faites pour déterminer quelles seraient les tailles optimales des grappes pour les moyennes et les proportions globales et trois valeurs différentes de la corrélation intragrappe, .005, .01 et .02, en utilisant les données sur les coûts tirées de l'enquête de production. (Ces résultats sont comparables à ceux qui ont été obtenus pour les variables d'attitudes politiques et d'intentions de vote dans les enquêtes réelles.) Les estimations des coûts C_a et C_b pour chaque strate figurent également dans le tableau. Le tableau montre que c'est dans la strate à faible densité que la taille optimale des grappes est la plus grande, ce qui reflète le fait que les coûts de filtrage sont élevés dans cette strate. À noter aussi que ces tailles des grappes optimales tendent à être plus grandes que celles qui ont effectivement été utilisées dans l'enquête (soit $b = 6.5$).

À noter encore que la taille optimale des grappes est à peu près la même pour les grappes tirées des UPE de 100 comme de 200 numéros consécutifs et que la perte d'efficacité en termes de coûts par suite de l'utilisation des séries de 200 plutôt que de 100 numéros consécutifs est minime. (Les estimations de la variance d'échantillonnage signifient également que la corrélation intragrappe est à peu près la même, indépendamment du fait que les grappes sont tirées des séries de 100 ou de 200 numéros consécutifs.)

Les valeurs de la taille optimale des grappes figurant dans le tableau 5 excèdent en général la valeur de la taille des grappes que permettent d'obtenir des UPE de 100 numéros consécutifs. Autrement dit, une proportion élevée d'UPE de 100 numéros consécutifs ne contiendrait pas suffisamment de numéros de ménages noirs pour produire des grappes de la taille désignée au second degré. Cette seule raison suffit à faire préférer les séries de 200 numéros consécutifs. Même avec les séries de 200 numéros consécutifs, il n'a pas été possible d'obtenir les grappes de la taille souhaitée au second degré pour toutes les UPE de la strate à faible densité. (Cela donne à penser que la vraie solution au problème du choix de la taille optimale des grappes devrait tenir compte de la capacité des UPE de produire des grappes; par ailleurs, la méthode utilisée ici peut certainement aider à prendre des décisions pratiques au sujet de l'efficacité en termes de coûts, mais elle ne prend pas en considération certaines situations extrêmes.)

Tableau 5

Paramètres de coûts et nombre optimal d'éléments par grappe selon la strate pour les grappes tirées des séries de 100 et de 200 numéros consécutifs et différentes valeurs de ρ

Strate et définition des grappes	Taille optimale des grappes			Paramètres de coûts	
	$\rho = .005$	$\rho = .01$	$\rho = .02$	C_{ha}	C_{hb}
Strate à forte densité					
100	15.9	11.2	7.9	\$50.81	\$40.11
200	15.9	11.2	7.9		
Strate à densité moyenne					
100	22.4	15.8	11.1	\$114.09	\$45.18
200	21.3	15.0	10.6		
Strate à faible densité					
100	29.8	21.0	14.8	\$309.98	\$69.52
200	29.9	21.1	14.8		

La deuxième décision importante qu'il nous a fallu prendre au sujet du plan de sondage avait trait au choix de la façon de répartir les unités de l'échantillon entre les strates. L'enquête utilisait des fractions de sondage dans des proportions de 3, 2 et 1 de la strate à forte densité à la strate à faible densité. Nous avons cherché à déterminer quelle serait la répartition optimale de l'échantillon entre les strates en supposant que les grappes choisies dans chaque strate était de taille optimale (voir tableau 5). Étant donné une taille fixée des grappes dans chaque strate, b_h , nous posons que la fraction de sondage dans la h -ième strate, f_h , est proportionnelle à $\sqrt{(Deff_h S_h^2)/(C_{ha}/b_h)}$, où $Deff_h$ est l'effet du plan de sondage pour la statistique dans la h -ième strate, S_h^2 est la variance de l'échantillon dans la h -ième strate, C_{ha} correspond aux coûts d'échantillonnage et de filtrage des UPE dans la h -ième strate et b_h est le nombre d'unités d'échantillon par grappe dans la h -ième strate.

Le tableau 6 présente les proportions optimales entre les fractions de sondage pour différentes combinaisons de variances dans chacune des trois strates et diverses valeurs de ρ . Le tableau montre que la répartition optimale de l'échantillon entre les strates est relativement peu sensible aux variations de ρ (pour les valeurs de ρ susceptibles d'être utilisées dans le plan de sondage). Quand les deux strates à plus forte densité de ménages noirs ont une variance au moins égale à celle de la strate à faible densité, il est préférable de prélever proportionnellement plus d'unités de sondage dans ces strates. (Cela reflète les coûts beaucoup moins élevés dans ces strates.) Quand l'écart-type est dans des proportions de 1.7, 1.5 et 1 environ entre les trois strates, le meilleur rapport entre les fractions de sondage est de 3, 2 et 1. L'analyse des données obtenues à partir de l'enquête donne à penser que, pour beaucoup de variables, l'écart-type est à peu près dans des proportions de 1, 1 et 1 entre les trois strates. Dans le cas de ces variables, les fractions de sondage, pour être optimales, doivent avoir un rapport de 1.7, 1.4 et 1 pour les tailles optimales des grappes présentées dans le tableau 5. (Pour une taille de grappe de 6.5, soit la taille effectivement utilisée dans chaque strate,

Tableau 6

Répartition optimale de l'échantillon entre les strates pour les moyennes globales, compte tenu de la formation de grappes de taille optimale dans chaque strate, pour diverses ratios des écart-types à l'intérieur des strates et diverses valeurs de la corrélation intragrappe

Ratios des écarts-types à l'intérieur des strates (Forte, moyenne et faible densité)	Ratios des fractions de sondage optimales (Forte, moyenne et faible densité)
$\rho = .005$	
3 : 2 : 1 1.7 : 1.5 : 1 1 : 1 : 1 .33 : .5 : 1	5.2 : 2.7 : 1 3 : 2 : 1 1.7 : 1.4 : 1 .6 : .9 : 1
$\rho = .01$	
3 : 2 : 1 1.7 : 1.5 : 1 1 : 1 : 1 .33 : .5 : 1	5.2 : 2.7 : 1 3 : 2 : 1 1.7 : 1.4 : 1 .6 : .9 : 1
$\rho = .02$	
3 : 2 : 1 1.8 : 1.5 : 1 1 : 1 : 1 .33 : .5 : 1	5.1 : 2.7 : 1 3 : 2 : 1 1.7 : 1.3 : 1 .6 : .9 : 1

les fractions de sondage optimales sont dans des proportions de 2.5, 1.6 et 1). D'après ces deux résultats, on peut dire que la méthode utilisée dans l'étude de production et qui consiste à prélever proportionnellement plus d'une unités d'une strate que d'une autre a entraîné, pour un coût unitaire donné, une perte de précision par rapport à la méthode correspondant à l'utilisation de fractions de sondage optimales.

6. SOMMAIRE

L'échantillonnage de populations rares oblige le statisticien qui veut s'en servir pour faire des enquêtes à prendre en considération diverses combinaisons de définitions des UPE et des grappes, de méthodes de stratification et de méthodes de mesure de la taille qu'on ne trouve pas habituellement dans les échantillons transversaux. Notre recherche nous a amenés à conclure qu'il est possible de modifier les techniques de sondage habituelles pour augmenter l'efficacité des échantillons à deux degrés pour les enquêtes téléphoniques auprès de la population des ménages noirs abonnés au téléphone.

Premièrement, nous avons constaté que même avec une correspondance très approximative entre les limites géographiques des circonscriptions téléphoniques et celles des grandes villes et des États, il a été possible de stratifier l'univers visé par l'enquête en strates qui permettaient de distinguer des groupes de circonscriptions possédant des taux d'admissibilité très différents. La proportion de ménages noirs était plus que deux fois plus élevée dans la strate à forte densité que dans la strate à faible densité. Cela permet

de réduire les coûts de filtrage au moment de l'élaboration du plan de sondage. Avec d'autres populations rares formées de groupes dont la concentration géographique est aussi nette que celle de la population noire des États-Unis, on peut s'attendre aux résultats à peu près analogues.

Deuxièmement, l'utilisation de règles de rejet fondées sur des critères d'admissibilité des unités d'une sous-population réduit effectivement les coûts de filtrage dans les UPE. Cela augmente la proportion admissible de numéros secondaires de deux à neuf fois selon la densité de la strate considérée.

Troisièmement, on ne perd pas beaucoup de numéros admissibles en utilisant des UPE plus grandes (c'est-à-dire des UPE de 200 plutôt que de 100 numéros consécutifs). Les séries de 100 à forte densité de numéros admissibles ont tendance à précéder ou à suivre d'autres séries à forte densité. Cela est un phénomène que nous n'avons pas encore observé concernant l'attribution des numéros par les compagnies de téléphone. Ce fait permet de prélever plus de numéros par UPE, ce qui est un autre élément clé de la réduction des coûts de production de l'échantillon de la population noire.

Malgré qu'on soit fortement tenté de chercher davantage à réduire les coûts dans le cas des enquêtes par sondage auprès de populations rares, il importe d'établir explicitement un juste compromis entre les erreurs et les coûts au moment de décider du plan de sondage définitif. Le modèle de coûts et d'erreurs d'échantillonnage que nous avons adopté pour notre recherche donne à penser qu'il est préférable de ne pas répartir l'échantillon proportionnellement entre les strates de population noire et qu'il est plus efficace de choisir un nombre relativement grand de numéros secondaires par UPE. Ce nombre devrait être suffisamment grand pour obliger à utiliser des UPE de 200 ou de 400 numéros consécutifs.

Même si nous n'avons appliqué ce plan de sondage qu'à la population noire, les résultats qu'il donne devraient être à peu près les mêmes pour d'autres populations à concentration géographique nette. Ces autres populations incluent notamment les groupes selon le revenu, certaines catégories professionnelles et les groupes ethniques.

En outre, les résultats auxquels notre recherche a abouti peuvent également avoir des répercussions sur les méthodes d'échantillonnage transversal. Il se peut qu'il soit avantageux d'augmenter la taille des UPE de 100 à 200 numéros consécutifs dans un plan de sondage par CNH à deux degrés qu'on voudrait appliquer à l'ensemble de la population des ménages abonnés au téléphone. Des séries plus grandes de 200 numéros consécutifs donneraient deux fois plus de numéros parmi lesquels choisir et, comme dans le cas d'enquêtes par sondage auprès de populations rares, la proportion de numéros admissibles aurait tendance à être la même que celle qu'on obtient avec des séries de 100 numéros consécutifs. Par conséquent, avec des valeurs faibles de la corrélation intragrappe, on pourrait décider d'avoir, pour un plan donné, un nombre de numéros admissibles par grappe qui serait beaucoup plus proche de la taille optimale. Parce que toutes les UPE choisies permettraient d'obtenir le nombre choisi de numéros admissibles, la distribution des grappes selon leur taille devrait aussi être moins variable sur l'ensemble des UPE et toute pondération compensatrice devrait donc avoir peu d'effet sur la variance des estimations.

BIBLIOGRAPHIE

- BLAIR, J., et R. CZAJA, (1982). Locating a special population using random digit dialing. *Public Opinion Quarterly*, 46, 585-590.
- GROVES, R.M., An empirical comparison of two telephone sample designs. *Journal of Marketing Research*, 15, 622-631.
- GROVES, R.M. et R.L. KAHN (1979). *Surveys by Telephone*. New York: Academic Press.

- KISH, L. (1976). Optima and proxima in linear sample designs. *Journal of the Royal Statistical Society*, Sér. A, 139, 80-95.
- KISH, L. (1965). *Survey Sampling*. New York: John Wiley.
- THORNBERRY, O.T., et J.T. MASSEY (1983). Coverage and response in random digit dialed national surveys. *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 654-659.
- WAKSBERG, J. (1978). Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 73, 40-46.